QCD & QMC

Developments in perturbative Quantum Chromo-Dynamics and Quasi Monte-Carlo

Een wetenschappelijke proeve op het gebied van Theoretische Hoge Energie Fysica

Proefschrift
ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus, Prof. dr. C.W.P.M. Blom
volgens besluit van het College van Decanen

in het openbaar te verdedigen op maandag, 26 maart 2007
om 13u30 stipt

door

Achillefs Lazopoulos

geboren op 13 november 1975 te Athene

Promotor:

prof.dr. R.H.P. Kleiss (IMAPP, RU Nijmegen)

Co-promotor:

dr. C.G. Papadopoulos (Institute of Nuclear Physics - NCSR Demokritos)

Names of the members of the Manuscript Committee:

Prof. Dr. S. de Jong (IMAPP, RU Nijmegen)
Prof. Dr. E. Laenen (NIKHEF, Amsterdam and Univ. Utrecht)
Dr. J. Vermaseren (NIKHEF, Amsterdam)
Dr. A. van Hameren (Inst. of Nucl. Physics, IFJ-PAM, Univ of Krakow, Poland)
Dr. P. Draggiotis (Univ. of Athens)

QCD & QMC

Developments in perturbative Quantum Chromo-Dynamics and Quasi Monte-Carlo.

An academic essay in Theoretical High Energy Physics

Doctoral thesis
to obtain the degree of doctor
from Radboud University Nijmegen
on the authority of the Rector Magnificus, Prof. dr. C.W.P.M. Blom,
according to the decision of the Council of Deans

to be defended in public on Monday, 26 March 2007
at 1.30 p.m. precicely

by

Achillefs Lazopoulos

born in Athens on November 13th, 1975

Promotor:

prof.dr. R.H.P. Kleiss (IMAPP, RU Nijmegen)

Co-promotor:

dr. C.G. Papadopoulos (Institute of Nuclear Physics - NCSR Demokritos)

Names of the members of the Manuscript Committee:

Prof. Dr. S. de Jong (IMAPP, RU Nijmegen)
Prof. Dr. E. Laenen (NIKHEF, Amsterdam and Univ. Utrecht)
Dr. J. Vermaseren (NIKHEF, Amsterdam)
Dr. A. van Hameren (Inst. of Nucl. Physics, IFJ-PAM, Univ of Krakow, Poland)
Dr. P. Draggiotis (Univ. of Athens)

Contents

List of Figures

Preface

Preface

Theoretical physics is a discipline whose scope and limitations are hard to define. However, in every scientific era, from Galileo till today, it is easy to recognize a compact front of knowledge that advances (slowly and with uneven steps), side by side with the ever increasing level of detail with which our experimental equipment probes the guts of nature. This front is currently located at the level of the sub-nucleonic regime, or, in the equivalent language of energies, at the TeV scale, and will certainly advance a bit further when, in late 2007, the Large Hadron Collider (LHC) is switched on. Those working on the front are, in a sense, soldiers of the current paradigm. This paradigm manifests itself, for the time being, through what is called 'The Standard Model'.

The Standard Model asserts that the world is built out of combinations of six quarks (and their anti-quarks) and six leptons (and their anti-leptons) interacting with each other with the mediation of the photon, the eight gluons and the three electroweak particles (called W^+,W^- and Z). The masses of these particles are generated in the presence of a yet speculative scalar particle, the Higgs particle, through the mechanism of spontaneous symmetry braking. The discovery of the Higgs is currently the holy grail of experimental high energy physics, and the main motivation for the commissioning of LHC.

The Standard Model describes how these particles interact with each other, within the counter-intuitive[1] framework of quantum field theory. It, thereby, introduces 19 constant parameters[2] the values of which have to be determined by experiment.

Table 1: The Standard Model: Leptons

symbol	name	mass (GeV/c^2)	charge
ν_e	electron neutrino	$< 10^{-8}$	0
e	electron	0.000511	-1
ν_μ	muon neutrino	< 0.0002	0
μ	muon	0.106	-1
ν_τ	tau neutrino	< 0.02	0
τ	tau	1.7771	-1

Producing exact quantitative results within the framework of quantum field theory is notoriously difficult. In most cases we are forced to seek information from the high-energy regime of the theory, where the interactions are weak enough to be treated as perturbations around a configuration where all particles are moving freely.

The popularity of such a perturbative approach to quantum field theory is not unrelated to the fact that our experimental configurations rely, since the heroic era of Rutherford, on high-energy particles colliding with each other. In such scattering experiments, the way the scattered particles are statistically distributed depends strongly (and in full detail) on the nature of the interaction that took place during the actual scattering. That distribution, measured by cross-sections, is abundant with information. Often there is far too much information for our taste!

[1] Counter-intuitive is a word that can be attached to anything in the quantum world. Of course, intuition is a matter of education, and one of the struggles of a Ph.D. student in High Energy Physics is to develop a kind of intuition within which the quantum phenomena do not appear to be so wonderfully magical any more.

[2] Excluding, that is, the 10 parameters further needed for the neutrino mass matrix.

Table 2: The Standard Model: Quarks

symbol	name	Approx. mass (GeV/c^2)	charge
u	up	0.003	2/3
d	down	0.006	$-1/3$
c	charm	1.3	2/3
s	strange	0.1	$-1/3$
t	top	174.2	2/3
b	bottom	4.3	$-1/3$

Table 3: The Standard Model: Electroweak bosons

name	mass (GeV/c^2)	charge
γ	0	0
W^+	80.4	+1
W^-	80.4	-1
Z	91.187	0

Table 4: The Standard Model: Strong Interaction bosons

name	Approx. mass (GeV/c^2)	charge
g	0	0

The LHC, the new generation particle collider, has been described repeatedly as a garbage bag collision theater. Protons will collide in the center of its four detectors after having been accelerated in the LHC ring to energies up to 7 TeV. Protons, however, are not elementary particles. They consist of a sea of quarks that emerge and disappear within a bewildering web of gluons, leaving three quarks always prevalent. The high energy collisions anticipated in LHC will explode these 'garbage bags' releasing all kind of particles that might themselves decay before detected, or disguise themselves by forming conglomerations of hadrons[3]. A small fraction of the detected particles will be recognized as potentially interesting: they might have

[3]Examples of hadrons include the proton and the neutron that constitute all nuclei, as well as a wild variety of quark composites that can be charged or neutral, of fermionic or bosonic nature, but in all cases having neutral color charge.

been produced by a Higgs particle, or by some new, yet undetected, beyond the Standard Model, candidate.

However, for every process that involves a Higgs, there are processes that don't and produce exactly the same products. Events in which these processes materialize are unwanted, 'background' noise that we need to subtract from our analysis. Furthermore, in most cases this 'background signal', in the language of experimentalists, will be overwhelmingly dominating the truly interesting events, the 'signal'. Most of it will involve particles that interact through the strong interaction. It will be a monstrous job of dust-cleaning an enormous amount of data before the interesting signal is extracted.

Table 5: The Large Hadron Collider

Particles used	Protons and Heavy Ions (Lead, fully stripped,82+)
E_{CM}	$14\ TeV$
Luminosity	$0.12 - 2.3 \times 10^{34}\ cm^{-2}s^{-1}$
Nominal energy for Protons	$7\ TeV$
Circumference of the accelerator ring	$26,659m$
Revolution frequency	$11.2455\ kHz.$
Operating temperature	$1.9\ K$
Magnetic field at 7 TeV	8.33 Tesla

However, what is the undesirable background signal of the present, has been the excruciatingly scrutinized, holy grail of the past. We know, in principle, all there is to know[4] about these undesirable out-coming particles: they are quarks and gluons, obeying the rules of Quantum Chromodynamics. The problem that arises is that now such particles can be produced in multitudes since the available energy is large enough. In other words, we need to know more about the final state distribution of particles in the case many of them (more than two and up to ten) are produced in a single collision.

In this thesis my work in three quite different problems within the field of multi-particle perturbative calculations is presented. At first, there is the problem of automating (and implementing in computer software) the calculation of the next-to-leading order term in the perturbation series of multi-particle cross-sections, and the consequent need to evaluate the complexity of different approaches therein. In chapter 2 I will present a quantitative description of the computational complexity of a straightforward diagram by diagram approach to the calculation of the one- and two- loop corrections in various theories, as opposed to the computational complexity of a recursive method assuming its extension beyond the tree level in a straightforward way.

Secondly, there is the problem of integrating multi-particle differential cross-sections numerically. In chapter 3 I seek to advance the use of an efficient but undervalued method for numerical integration, the method of Quasi-Monte Carlo, by attempting to remedy the main obstruction to its use: the lack of a reliable statistical estimate for the induced error. Such an estimate is presented together with a number of alternative approaches that trade accuracy for computational complexity.

Finally, there is the problem of obtaining analytical expressions for the leading-order expansion of multi-particle cross-sections. Among the recent developments that have taken place in the last few years, a new type of recursion relations for tree-level QCD amplitudes has appeared. In chapter 4 I present a direct proof of that recursion relation in terms of a diagrammatic correspondence with ordinary Feynman diagrams, that sheds light on the central role played by

[4] In principle, since the precise way in which quarks and gluons hadronize still escapes a quantitative description.

gauge invariance in this recursive construction.

The organization of this thesis follows this three-fold schema, preceded by an introductory chapter where a number of prerequisite topics are presented. The two appendices at the end contain some information on Monte-Carlo estimators and on implementing algorithms for constructing Quasi-Monte Carlo pointsets, that the Monte-Carlo practitioner might find of use. The plots appearing in chapter 3 of this thesis, an implementation of the Quasi-Monte Carlo error estimate as well as of the code needed to produce the Niederreiter sequences of the appendix can be found at `http://www.ru.nl/imapp/theory/mc_qmc`.

Chapter 1

Introduction

In this chapter we will introduce, for clarity, elementary concepts of field theory and Quantum Chromo-Dynamics, reciting formulas like that for the cross-section and the decay rate, the Schwinger-Dyson equation in zero dimensions, the QCD Lagrangian and the Feynman rules for QCD. None of the topics here is, of course, original. Most of them can be found in textbooks like [37], [56], [59], [67] and many others.

Notation

In this thesis we will always use the diagonal Minkowski metric, with

$$g^{00} = 1,\ g^{ii} = -1 \tag{1.1}$$

Middle alphabet Greek letters $(\mu, \nu, \ldots)$ will be used for Lorentz space-time indices (unless otherwise specified), running from 0 to 3, whereas middle alphabet Roman letters $(i, j, k, \ldots)$ will be used for space indices, running from 1 to 3. Color indices in the adjoint representation will be denoted with Roman letters from the beginning of the alphabet $(a, b, c, \ldots)$, whereas in the fundamental representation by middle alphabet Roman ones $(i, j, k, \ldots)$. There will be a summation implied whenever two identical indices appear in the same monomial, according to the usual Einstein convention.

1.1 Cross-sections and decay widths

The main measurable quantities in modern high-energy scattering experiments are the differential cross-section and the differential decay width, or quantities derived from them. In this section we present, in short, the main steps in the derivation of the formula for the cross-section and the decay width for a given scattering process.

The differential cross-section is the transition probability of a system to go from an initial state i to a final state f, per scatterer in the target per unit flux of the incident beam.

$$\frac{d\sigma}{\prod_f [dp_f]} = \frac{W_{i\to f}}{N_{scat.} \times F_{inc.}} \tag{1.2}$$

The initial state is usually considered to consist of two particles with well-defined momenta, far enough from each other to be thought of as free from interactions. In reality, two beams of particles are prepared, in which particles have momenta spread as narrowly as possible around a mean value. The incident flux is defined as the number of particles in one of the beams that cross a unit area in the rest frame of the other beam, per second.

Let us assume that the particles in the two beams have momenta p_1 and p_2 respectively. The initial state of the scattering is then modeled by two particles of momenta p_1 and p_2. The denominator in eq.(1.2) provides a so-called kinematical factor of $\frac{1}{2E_2 2|\vec{p}_1|}$, calculated in the rest frame of particle 2. The numerator contains all the dynamical information about the scattering provided by the theory. It contributes $2\pi\delta^4\left(p_1 + p_2 - \sum p_f\right)|A_{1,2\to f}|^2$, where $|A_{1,2\to f}|^2$ is the quantum amplitude for the transition from the initial to the final state that depends on the model and the particles involved. The final result for the differential cross-section is

$$\frac{d\sigma}{\prod_f [dp_f]} = \frac{dW_{i\to f}}{d^4x \frac{dN_{sct.}}{d^4x} \times F_{incident}} = \frac{1}{2E_2 2|\vec{p}_1|}(2\pi)^4\delta^4(p_1 + p_2 - \sum p_f)|A_{1,2\to f}|^2 \tag{1.3}$$

or

$$\frac{d\sigma}{\prod_f [dp_f]} = \frac{1}{2E_2 2|\vec{p}_1|}(2\pi)^4\delta^4(p_1 + p_2 - \sum p_f)|A_{1,2\to f}|^2 \tag{1.4}$$

Note that in the rest frame of particle 2 we have $E_2 = m_2$. One could write the kinematical factor in a Lorentz-invariant way

$$2E_2 2|\vec{p}_1| = 4\sqrt{(p_1 \cdot p_2)^2 - m_1^2 m_2^2} \tag{1.5}$$

In any rest frame where the initial particles are collinear, so that $(\vec{p}_1 \cdot \vec{p}_2)^2 = |\vec{p}_1|^2|\vec{p}_2|^2$, the above expression is equivalent with the more frequently used

$$2E_2 2|\vec{p}_1| = 2E_1 2E_2|v_1 - v_2| \qquad v_i \equiv \frac{|\vec{p}_i|}{E_i} \qquad \vec{p}_1 \| \vec{p}_2 \tag{1.6}$$

So we have

$$\frac{d\sigma}{\prod_f [dp_f]} = \frac{1}{2E_1 2E_2|v_1 - v_2|}(2\pi)^4\delta^4(p_1 + p_2 - \sum p_f)|A_{1,2\to f}|^2 \tag{1.7}$$

The decay width of a particle that decays to a number of debris is defined to be the number of decays per unit time per decaying particle

$$\frac{d\Gamma_{i\to f}}{\prod_f [dp_f]} = \frac{dW_{i\to f}}{dN_{dec.}} \tag{1.8}$$

where, as before, $dN_{dec.}/d^4x = 2E|f(x)|^2 = 2m|f(x)|^2$ and the above reasoning is repeated to arrive at

$$\frac{d\Gamma}{\prod_f [dp_f]} = \frac{1}{2m}(2\pi)^4\delta^4(p_1 + p_2 - \sum p_f)|A_{1\to f}|^2 \tag{1.9}$$

To find total cross-sections or decay widths one has to integrate eq.(1.7) or eq.(1.9) over the final particles' phase-space[1]:

$$\sigma = \int \prod_f [dp_f] \frac{1}{2E_1 2E_2 |v_1 - v_2|} (2\pi)^4 \delta^4(p_1 + p_2 - \sum p_f) |A_{1,2\to f}|^2 \tag{1.10}$$

$$\Gamma = \int \prod_f [dp_f] \frac{1}{2m} (2\pi)^4 \delta^4(p_1 + p_2 - \sum p_f) |A_{1\to f}|^2 \tag{1.11}$$

where the measure for each final state momentum is

$$[dp_f] = \frac{d^4 p_i}{(2\pi)^4} \delta(p_i^2 - m^2) \theta(p_i^0) \tag{1.12}$$

Here, the δ-function imposes the on-shell condition for every outgoing particle and the θ-function imposes the demand that only positive energy solutions contribute to observables. It is evident that the phase-space integral has dimensionality equal to $3N - 4$ (four degrees of freedom are subtracted due to overall momentum conservation, imposed by the four-dimensional δ-function in eq.(1.10,1.11).

In any process involving more than three final state particles, the phase-space integral cannot be performed analytically. One has to resort to numerical methods in order to reach a result that can be compared with experimental data. We will see in chapter 3 that such numerical methods have been implemented in a number of computer programs, in various degrees of efficiency and generality.

[1] If the desired cross-section or decay rate refer to particles of which the spin, helicity, color or flavor are not detected, one has to sum over all these discrete degrees of freedom of the final state as well.

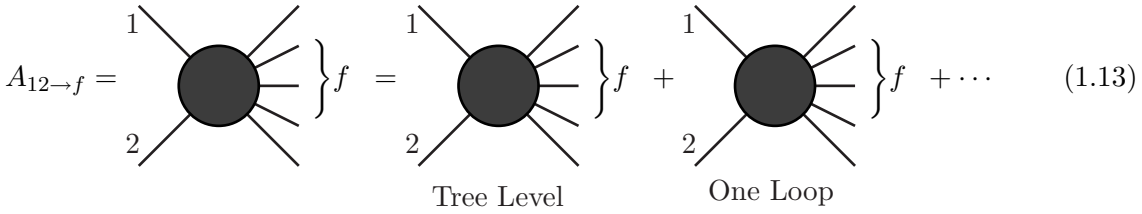
1.2 The amplitude

The computationally expensive part of cross-section or decay width calculations is always the evaluation of the amplitude, $A_{1,2\to f}$ in eq.(1.10),or $A_{1\to f}$ in eq.(1.11). The amplitude is the quantity that carries all the dynamical information of the scattering process, thus revealing the nature of the interactions between the involved particles.

At present, we can only evaluate the amplitude as a perturbation series over the coupling constants of the theory that models the interaction. The convergence of such a series is an entirely non-trivial matter[2], but its first few terms approximate the value of the amplitude very well when the coupling constants (used as the expansion parameters) are small enough. In Quantum Chromodynamics, this is achieved in relatively high energies, due to a phenomenon called 'asymptotic freedom'.

Each term in the perturbation series is itself a sum of quantities conveniently encoded in 'Feynman diagrams'. Feynman diagrams are graphs[3] made out of lines and vertices. The type of lines and vertices allowed depends on the type of particles involved in the process we study. This information, as well as the precise Feynman rules for transcribing diagrams into algebraic quantities, is encoded in the 'Lagrangian density' of the theory (a function of the fields corresponding to the particles involved) that lies in the heart of every quantum field theory model.

Calculating the amplitude amounts to drawing all the Feynman diagrams for the leading order of the series, transcribing the diagrams into algebraic quantities with the help of recipes called 'Feynman rules', adding them up, then proceeding to the next to leading order etc.[4] . In case there is only one coupling constant in the theory, the order of a term in the perturbation series coincides with the number of closed loops in each diagram contributing to that term. The first term in the series, containing no loops, is called the 'tree level'.



(1.13)

The dark blobs denote the sum of all topologically connected and amputated[5] diagrams.

When the final state contains many outgoing particles, calculating by hand the mere tree-level contribution to the amplitude, by evaluating every single diagram, can be an impossible task[6]. Since the number of diagrams increases roughly as $N!$, with N the total number of external legs, in cases with $N \geq 8$, the calculation can be too intense even for a computer evaluating the diagrams numerically. It is much preferable to employ recursive techniques based on the Schwinger-Dyson equation, presented in the following section.

Would it be preferable to employ recursive techniques also in the one- and two-loop level? This is the topic that will concern us in chapter 2.

---

[2] Actually the series is known to diverge after a - sufficiently large - number of terms, and is always treated as an asymptotic series.

[3] In mathematics, the notion of a graph is wider than the one used here, but Feynman diagrams are graphs in the mathematical sense as well. Mathematicians use the word 'edge' for what we call 'line'.

[4] Naturally, there is a high degree of sophistication involved in deriving the following guidelines from first principles. The gory details can be found in all textbooks on quantum field theory, see for example [37], [56], [72], [67].

[5] In the sense that they don't contain self-energy subgraphs on the external legs. Such terms can be absorbed in the 'renormalization' constants of the external fields, i.e. in the definition of the asymptotic states. Any such subgraph is, thus, removed, and the graph that remains is termed 'amputated'.

[6] Though the situation has improved thanks to recent developments which will be presented in chapter 4.

1.3 The Schwinger-Dyson equations.

In the path-integral formalism, the main quantity of interest is the generating functional[7]

$$Z[J] = \int D\varphi \; e^{i\int d^4x \; \mathcal{L}(\varphi(x))+J(x)\varphi(x)} \tag{1.14}$$

Knowledge of the generating functional yields immediately any n-point function[8], by repeated functional differentiation:

$$G_n(x_1, x_2, \ldots, x_n) \equiv \langle 0|\mathcal{T}\varphi(x_1)\ldots\varphi(x_n)|0\rangle = \frac{1}{Z_0}(-i\frac{\delta}{\delta J(x_1)})\ldots(-i\frac{\delta}{\delta J(x_n)})Z[j]|_{J=0} \tag{1.15}$$

where $Z_0 = Z[J=0]$ and the functional differentiation is defined by

$$\frac{\delta}{\delta J(y)}J(x) = \delta^4(x-y) \tag{1.16}$$

When the functional differential operator acts on derivatives of its argument one has to perform integration by parts on the differentiated quantity to remove the derivative

$$\frac{\delta}{\delta\varphi(y)}\int d^4x \vartheta_\mu\varphi(x)Q^\mu(x) = -\frac{\delta}{\delta\varphi(y)}\int d^4x\varphi(x)\vartheta_\mu Q^\mu(x) = \vartheta_\mu Q^\mu(y) \tag{1.17}$$

The Schwinger-Dyson equations are 'equations of motion' for the Green's functions of the theory. At the level of the generating functional, they can be derived from

$$0 = \int D\varphi \; \frac{\delta}{\delta\varphi(y)} e^{i\int d^4x \; \mathcal{L}(\varphi(x))+J(x)\varphi(x)} \tag{1.18}$$

which can be seen as the demand that the path integral doesn't change under an infinitesimal shift of the fields $\varphi(x) \to \varphi(x) + \epsilon(x)$. Such a change would produce the term on the right hand side of eq.(1.18), which should, therefore, vanish. We have

$$\begin{aligned}
0 &= \int D\varphi \; \frac{\delta}{\delta\varphi(y)}\left[i\int d^4x \; \mathcal{L}(\varphi(x)) + J(x)\varphi(x)\right] e^{i\int d^4x \; \mathcal{L}(\varphi(x))+J(x)\varphi(x)} && (1.19)\\
&= \int D\varphi \; i\left[\frac{\vartheta\mathcal{L}}{\vartheta\varphi(y)} - \vartheta_\mu\frac{\vartheta\mathcal{L}}{\vartheta_\mu\varphi(y)} + J(y)\right] e^{i\int d^4x \; \mathcal{L}(\varphi(x))+J(x)\varphi(x)} && (1.20)\\
&= i\left[\frac{\vartheta\mathcal{L}}{\vartheta\varphi(y)} - \vartheta_\mu\frac{\vartheta\mathcal{L}}{\vartheta_\mu\varphi(y)} + J(y)\right]_{\varphi(y)=-i\frac{\delta}{\delta J(y)}} \int D\varphi \; e^{i\int d^4x \; \mathcal{L}(\varphi(x))+J(x)\varphi(x)} && (1.21)\\
&= i\left[\frac{\vartheta\mathcal{L}}{\vartheta\varphi(y)} - \vartheta_\mu\frac{\vartheta\mathcal{L}}{\vartheta_\mu\varphi(y)} + J(y)\right]_{\varphi(y)=-i\frac{\delta}{\delta J(y)}} Z[J] && (1.22)
\end{aligned}$$

or

$$JZ[J] = -\left[\frac{\vartheta\mathcal{L}}{\vartheta\varphi} - \vartheta_\mu\frac{\vartheta\mathcal{L}}{\vartheta_\mu\varphi}\right]_{\varphi=-i\frac{\delta}{\delta J}} Z[J] \tag{1.23}$$

The alarmingly formal level of eq.(1.23) can be made concrete when one considers the extreme, but very useful limit in which the dimensionality of the theory is reduced to zero[9]. In zero dimensions space-time consists of just one point, a field is a random variable that can take real values and the path integral is an ordinary integral over that random variable. The action is

[7] In order to keep the discussion simple we will use a scalar theory as a model theory in this section.

[8] Hence, also, any n-particle scattering amplitude, via the LSZ reduction formula.

[9] See [3] for details in zero-dimensional field theory

a function of the random variable, instead of a functional, and it contains no kinetic terms[10]. Hence, we have

$$Z(J) = \int_{-\infty}^{\infty} d\varphi e^{-S(\varphi)+J\varphi} \tag{1.24}$$

and the Schwinger-Dyson equation becomes

$$JZ(J) = \left[S'(\varphi)\right]_{\varphi=\frac{d}{dJ}} Z(J) \tag{1.25}$$

In the most general case, $S(\varphi) = \frac{\mu}{2}\varphi^2 - V(\varphi)$, with $V(\varphi) = \sum_{n=3}^{\infty} \frac{\lambda_n}{n!}\varphi^n$. Then we have

$$JZ(J) = \left[\mu\varphi - V'(\varphi)\right]_{\varphi=\frac{d}{dJ}} Z(J) = \mu\frac{d}{dJ}Z(J) - V'(\frac{d}{dJ})Z(J) \tag{1.26}$$

One can define the generating function for the connected Green's functions

$$W(J) = \log(\frac{Z(J)}{Z(0)}) \tag{1.27}$$

and the so-called amplitude-generating function

$$\phi(J) = \frac{dW}{dJ} = \frac{1}{Z(J)}\frac{dZ}{dJ} \tag{1.28}$$

The Schwinger-Dyson equation for this function is found by dividing eq.(1.26) by $Z(J)$:

$$\phi(J) = \frac{J}{\mu} + \frac{1}{Z(J)}V'(\frac{d}{dJ})Z(J) \tag{1.29}$$

The zero-dimensional Schwinger-Dyson equation is very convenient when one wishes to count diagrams. Setting the contribution of each Feynman diagram equal to 1 is fairly easy: it amounts to setting all propagators, vertices and external legs equal to 1 in the Feynman rules. Then an amplitude with n external legs would be equal to the number of diagrams with n legs, since each diagram equals 1, and counting is 'reduced' to calculating the relevant amplitudes. In zero dimensions, this can be achieved by simply setting the coupling constants λ_n and the 'mass' μ to 1. The amplitude-generating function $\phi(j)$ is, then the generating function for the numbers $a_{1\to n}$ of connected diagrams with $n+1$ external legs.

$$\phi(J) = \sum_{n=0}^{\infty} \frac{a_{1\to n}}{n!} J^n \tag{1.30}$$

In this context, it can be seen that derivatives of $\phi(J)$ in the Schwinger-Dyson equation for $\phi(J)$, eq.(1.29), actually count diagrams with loops[11]. When one is interested in counting tree-level diagrams, therefore, one can ignore terms with derivatives of $\phi(J)$ in the Schwinger-Dyson equation arriving at

$$\phi(J) = J + V'(\phi(J)) \tag{1.31}$$

This equation will be put in use in chapter 2, where counting of diagrams in tree level and beyond is the main subject.

[10]Kinetic terms are a consequence of the way the value of the field at a space-time point influences the field values at neighboring points. In zero dimensions there are no such points!

[11]This can, alternatively, be seen by re-inserting the $\hbar$ parameter in eq.(1.24). Then $S \to \frac{1}{\hbar}S$ and derivative terms in eq.(1.29) carry higher order powers of $\hbar$.

1.4 Elements of QCD

Quantum Chromo-Dynamics (QCD) is the theory that describes the physics of strong interaction. The elementary particles that can interact strongly are the six quarks (u,d,s,c,t,b) and the gluon, all of them carrying color charge[12]. QCD is a non-Abelian gauge theory, also called a Yang-Mills theory. The meaning of this is that the theory is constructed to be invariant under a non-Abelian gauge symmetry group, in this case $SU(3)$.

1.4.1 The QCD Lagrangian before quantization

The classical Lagrangian density of QCD theory is

$$\mathcal{L}_{QCD} = \overline{\psi}(x)(i\gamma^\mu D_\mu - m)\psi(x) - \frac{1}{4}F^{\mu\nu,a}F^a_{\mu\nu} \tag{1.32}$$

where

$$D_\mu = \vartheta_\mu + igt^a A^a_\mu \,, \tag{1.33}$$

$$F^a_{\mu\nu} = \vartheta_\mu A^a_\nu - \vartheta_\nu A^a_\mu - gf^{abc}A^b_\mu A^c_\nu \,, \tag{1.34}$$

γ^μ are the four Dirac matrices, t^a_{ij} the 8 generators of the $SU(3)$ group ($a = 1 \ldots 8$) and f^{abc} the structure constants of the group.

It is worth noticing that the object $\psi(x)$ carries, apart from spinorial, color indices. It actually contains three spinors, denoted by $\psi_i(x)$, $i = 1, 2, 3$, each of which has four complex components. These color indices are to be contracted[13] with the color indices of the $SU(3)$ generators t^a_{ij}.

There is a variety of conventions on t^a matrices in the literature, and any variant involves always a choice on the normalization of the color generators of the $SU(3)$ group, and a potential choice in the definition of the structure constants. Here we use

$$Tr(t^a t^b) = \frac{1}{2}\delta^{ab} \tag{1.36}$$

$$[t^a, t^b] = if^{abc}t^c \tag{1.37}$$

which lead to

$$f^{abc} = -2i[Tr(t^a t^b t^c) - Tr(t^a t^c t^b)] \tag{1.38}$$

1.4.2 Gauge Symmetry and the quantized Lagrangian

The classical Lagrangian of QCD is (constructed to be) invariant under a 'gauge transformation'

$$\psi(x) \to \psi'(x) = U(\theta^a(x))\psi(x) \quad U(\theta^a(x)) \equiv e^{-igt^a\theta^a(x)} \tag{1.39}$$

It is easy to see that the Lagrangian remains invariant if

$$D_\mu \to D'_\mu = UD_\mu U^{-1} \tag{1.40}$$

which means

$$t^a A^a_\mu \to t^a A^{a\prime}_\mu = Ut^a A^a_\mu U^{-1} + \frac{i}{g}(\vartheta_\mu U)U^{-1} \tag{1.41}$$

[12] There are three kinds of color charge, as opposed to one kind in electromagnetism.

[13] Strictly speaking, then, eq.(1.33), for example, should have been written as

$$(D_\mu)_{ij} = \vartheta_\mu \delta_{ij} + igt^a_{ij}A^a_\mu \tag{1.35}$$

We have suppressed such color indices for simplicity.

Expanding the exponential of U in series and discarding terms of order higher than linear in g, in eq.(1.41), we get

$$A^a_\mu(x) \to A^{a\prime}_\mu(x) = A^a_\mu(x) + \vartheta_\mu \theta^a(x) + g f^{abc} \theta^b A^c_\mu \tag{1.42}$$

Since the Lagrangian is invariant under gauge transformations, field configurations that differ by such a gauge transformation lead to the same physics, and should be considered completely equivalent. This effect has an undesirable consequence when one tries to quantize the theory. In terms of canonical quantization, one is faced with the fact that there are too many degrees of freedom for the gluon field[14], which results to the vanishing of the canonical momentum conjugate to the zero component of the gluon field[15]. In terms of the path-integral approach, contributions from equivalent gluon-field configurations are counted separately, thus inflicting an infinite overcounting of contributions. Equivalently, one can see that the corresponding operator in the Euler-Lagrange equation of motion for the gluonic field is not invertible.

To circumvent this problem one could, in the case of an Abelian gauge theory (like QED - Quantum ElectroDynamics) use the method of Lagrange multipliers to constrain the equation of motion, adding to the Lagrangian a gauge-fixing term of the type $\mathcal{L}_{gaugefix.} = -\frac{(\vartheta_\mu A^\mu)^2}{2\xi}$, where ξ is a gauge-fixing parameter we expect to vanish in any physically measurable quantity. However, due to the non-Abelian nature of the gauge symmetry group, it is preferable to proceed through the path integral formalism and the Fadeev-Popov procedure, which introduces, apart from the above gauge-fixing term, an extra term involving 'ghost' fields, $\mathcal{L}_{ghost}$.

$$\mathcal{L}_{QCD} = \overline{\psi}(x)(i\gamma^\mu D_\mu - m)\psi(x) - \frac{1}{4}F^{\mu\nu,a}F^a_{\mu\nu} - \frac{(\vartheta\cdot A)^2}{2\xi} - \mathcal{L}_{ghost} \tag{1.43}$$

Following the Fadeev-Popov procedure, one realizes that the technical reason for the appearance of ghosts is the fact that the gauge-fixing term depends on the derivative of the gluon field. One can, therefore, avoid the ghosts alltogether[16], by using a different gauge-fixing term, like the one of the axial gauges $\mathcal{L}_{gaug.fix} = -\frac{(n_\mu A^{a,\mu})^2}{2\xi}$ at the limit $\xi \to 0$. The disadvantage here is the introduction of the special vector n^μ which should drop out of every physical result.

$$\mathcal{L}_{QCD} = \overline{\psi}(x)(i\gamma^\mu D_\mu - m)\psi(x) - \frac{1}{4}F^{\mu\nu,a}F^a_{\mu\nu} - \frac{(n_\mu A^{a,\mu})^2}{2\xi} \tag{1.44}$$

1.4.3 Gauge invariance

It is instructive to note that both gauge-fixing terms presented in sec. 1.4.2 induce terms in the gluon propagator that are proportional to k^μ, the momentum carried by the propagator. These contribute something non-zero in a Feynman diagram, but their contributions should vanish in the level of the amplitude. This is, of course, equivalent with the statement that the Ward-Takahashi identity holds in a non-Abelian theory, that is,

$$M_\mu(k,\ldots)k^\mu = 0 \tag{1.45}$$

for every 'current' M_μ even if k^μ is off-shell, provided that all the others are on-shell. This, in turn, is equivalent with current conservation, and can be traced back to the level of the Lagrangian, as an effect of local gauge invariance!

One could see gauge invariance from another point of view: not as a postulate but as a necessity. Unitarity arguments can be used to show that the non-transverse degree of freedom of any massless gauge boson should not contribute to physically measurable quantities. This implies the Ward-Takahashi identity. Furthermore, any theory where there are interactions

[14] Four instead of two.

[15] This is obviously a problem, since the canonical commutation relations can never be satisfied.

[16] See [37] for details.

that change the charge of a particle - as is the case with QCD or the Electroweak theory - needs bosonic mediators that carry charge[17], to maintain current conservation. Any charged boson particle can, of course, interact with itself via a three-vertex interaction. Moreover, unitarity arguments can be used again to show that a four-vertex interaction is necessary, with a coupling constant equal to the square of the three-vertex coupling constant. In this context, gauge invariance can be seen as a way to incorporate, at the Lagrangian level, the terms that guarantee all the above.

1.4.4 The Feynman rules

The Feynman rules for Quantum Chromo-Dynamics involve, apart from quarks and gluons, 'ghost' lines naturally arising from the Fadeev-Popov procedure, in a general, Feynman gauge (see eq.(1.43)). We will not use these ghost lines in the present thesis, but we recite the corresponding rules here for completeness.

The Feynman rules in the axial gauge are identical with these in the Feynman gauge, except the rule for the gluon propagator. We give, in table 1.4, the corresponding propagators. Note that the axial vector n^m can be taken massless, $n^2 = 0$, thereby canceling the last term in the axial gauge propagator.

Table 1.1: QCD Feynman rules: propagators

	$\delta_{ab}\left(-g_{\mu\nu}+(1-\xi)\frac{p_\mu p_\nu}{p^2}\right)\frac{i}{p^2}$	$a,b=1\ldots8$
	$\delta_{jk}\frac{i(\not{p}+m)}{p^2-m^2}$	$i,j=1\ldots3$
	$\delta_{ab}\frac{i}{p^2}$	$a,b=1\ldots8$

The last of these rules, involving the 'symmetry factor' diagram, will be of importance in chapter 2. 'Symmetry factor' is a number associated with every Feynman diagram that reflects its 'topological' symmetry, counting the ways in which one could interchange lines or vertices of the diagram without changing its topology. The complication of the symmetry factors arises from the way in which Feynman diagrams are derived via the Wick theorem and the corresponding combinatorics. There, contributions corresponding to each diagram are counted with a 'default' combinatorial factor. Diagrams that have equivalent lines or vertices are overcounted and one has to multiply them by hand with the symmetry factor that reflects these symmetry properties[18]. In tree-level such symmetry factors are always equal to 1.

[17] As opposed to the chargeless photon.
[18] For details see [56]

Table 1.2: QCD Feynman rules: vertices

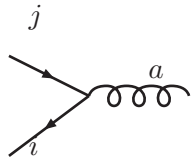
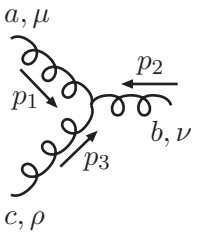
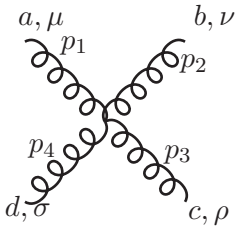
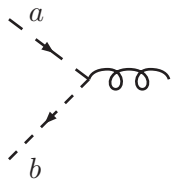
	$-igt^a_{ij}\gamma^\mu$	$a = 1 \dots 8$ $i, j = 1 \dots 3$
	$gf^{abc}[g_{\mu\nu}(p_1-p_2)_\rho + g_{\nu\rho}(p_2-p_3)_\mu + g_{\rho\mu}(p_3-p_1)_\nu]$	$a, b, c = 1 \dots 8$
	$-ig^2 f^{kac} f^{kbd}(g_{\mu\nu}g_{\rho\sigma} - g_{\mu\sigma}g_{\nu\rho})$ $-ig^2 f^{kad} f^{kbc}(g_{\mu\nu}g_{\rho\sigma} - g_{\mu\rho}g_{\nu\sigma})$ $-ig^2 f^{kab} f^{kcd}(g_{\mu\rho}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\rho})$	$a, b, c, d, k = 1 \dots 8$
	$gf^{abc}p^\mu$	$a, b, c = 1 \dots 8$

Table 1.3: Extra Feynman rules.

Moreover one has to

- integrate over any loop momentum k with a measure of $\int \frac{d^4k}{(2\pi)^4}$
- multiply by -1 for each quark or ghost loop
- multiply with the symmetry factor of each diagram

Table 1.4: Gluon propagators in different gauges

	$\delta_{ab}\left(-g_{\mu\nu} + (1-\xi)\frac{p_\mu p_\nu}{p^2}\right)\frac{i}{p^2}$	Feynman gauge
	$\delta_{ab}(-g_{\mu\nu} + \frac{p_\mu n_\nu + p_\nu n_\mu}{pn} - \frac{n^2 p_\mu p_\nu}{(pn)^2})\frac{1}{p^2}$	Axial gauge

Chapter 2

Counting Feynman diagrams

Methods for counting Feynman diagrams with loops for theories with bosonic particles and arbitrary vertices are presented. Diagrams are counted with symmetry factors and asymptotic results for one and two loops are given. The computational complexity of the Caravaglios-Moretti algorithm in the loop level with effective vertices is cited, and a comparison of that algorithm with the ordinary diagram per diagram approach is presented.

2.1 Introduction

Anticipating high-energy colliders of the caliber of LHC, high-multiplicity final states will become ever more relevant, increasing the need for efficient evaluation of complicated multi-leg amplitudes. Performing such calculations by a direct evaluation of all relevant Feynman graphs is computationally hard in the sense that the number of graphs increases with N - the total number of external legs - roughly as $N!$. For example, the $2 \to 8$ purely gluonic amplitude in QCD contains 10.5 million Feynman graphs at the tree level [25]; and one certainly expects that loop corrections (described by many more diagrams) will also be important.

A computational breakthrough has been achieved in the '80s by the introduction of the Caravaglios-Moretti (CM) algorithm [16], in which the Schwinger-Dyson (SD) equations of the theory are employed rather than their decomposition in individual Feynman diagrams. In this way, the full amplitude is evaluated in a recursive manner, thus achieving a complexity of order c^N (where c is a constant). Such methods, however, have, to date, only been formulated for the Born approximation. Barring a revolutionary new method for solving the SD equations including loop effects[1], the most straightforward approach would seem to use the vertices of the effective, rather than those of the bare, action. In such an approach the effective vertices incorporate contributions from any possible loop integral, and, therefore vertices with up to N legs have to be employed (instead of the three- and four- point vertices that are employed in a Yang-Mills type amplitude). This increases the complexity of the CM algorithm. In order to assess the relative merit of the CM algorithm, it is, therefore, relevant to compare the computational complexity of this 'effective' CM approach to the number of higher-order Feynman graphs.

In this chapter, we calculate the number of individual diagrams, not weighted by their symmetry factors, in zero-, one- and two-loop level for four models of a self-interacting scalar theory. We also give the number of one-particle-irreducible graphs, that is needed in the sequence. Moreover, we give the number of diagrams, now weighted by their symmetry factors, for the four models, as they occur directly from the path integral. We proceed with asymptotic estimates, in the number of external legs, for weighted and unweighted graphs. The computational complexity of the CM algorithm in one and two loops is calculated and a comparison of the efficiency of the CM algorithm to that of the individual-diagram approach is presented.

[1] In informal discussions, all the experts agree that this would be a tremendous advance — but no one has a clue on how to approach it.

2.2 Counting diagrams

We will consider a self-interacting scalar theory with arbitrary vertices of the type φ^k, $k = 3, 4, \ldots$. We define the 'potential'

$$V(\varphi) = \sum_{k\geq 3} \epsilon_k \, \frac{\varphi^k}{k!} \ , \tag{2.1}$$

where ϵ_k is 1 if the φ^k interaction is present, otherwise it is zero. We shall specialize to a number of cases:

$$\begin{aligned} \varphi^3 \text{ theory} &: V(\varphi) = \varphi^3/6 \ , \\ \varphi^4 \text{ theory} &: V(\varphi) = \varphi^4/24 \ , \\ \text{gluonic QCD} &: V(\varphi) = \varphi^3/6 + \varphi^4/24 \ , \\ \text{effective theory} &: V(\varphi) = e^{\varphi} - 1 - \varphi - \varphi^2/2 \ ; \end{aligned} \tag{2.2}$$

but alternative theories are easily implemented. Theories with fermions can in principle be considered without much difficulty, using multiple generating functions of many arguments, along the lines of [25].

2.2.1 A short note on generating functions

In general, when one needs to count configurations D_n of which the number, a_n depends on a discrete index n that labels objects of some type, it is convenient to use a generating function

$$G(x) = \sum_{n=1}^{\infty} \frac{x^n}{n!} a_n \tag{2.3}$$

Knowledge of the generating function in a closed form immediately yields all needed a_n through differentiation:

$$a_n = \left[\frac{d^n G(x)}{dx^n} \right]_{x=0} \tag{2.4}$$

Let us call A_n a collection of such configurations with the important property that it can be constructed by combining in all possible ways configurations that belong to collections A_t, A_s with $n = t + s$. In other words, the collection A_n contains all the possible configurations D_n, in each of which one can recognize configurations D_s and D_t that belong to A_s and A_t. Moreover, one can group the D_n in sub-collections, each consisting of merging configurations that belong to A_s and A_t, for all values of s and t. Then, given the numbers a_t, a_s one can easily find a_n as the number of ways to select s objects out of n:

$$a_n = \sum_s \frac{n!}{s!(n-s)!} a_s a_{n-s} \tag{2.5}$$

Note, however, that one has to divide by a factor of $\frac{1}{2}$ to compensate for the overcounting of cases like $t = k, s = n - k$ and $t = n - k, s = k$. Now, the square of the generating function is

$$G^2(x) = \sum_{t,s=1}^{\infty} \frac{x^{t+s}}{t!s!} a_t a_s \tag{2.6}$$

whose n'th derivative at zero is

$$\left[\frac{d^n G^2(x)}{dx^n} \right]_{x=0} = \sum_s \frac{n!}{s!(n-s)!} a_s a_{n-s} \tag{2.7}$$

Therefore, including the 'symmetry' factor $\frac{1}{2}$, $\frac{1}{2}G^2(x)$ is actually generating the correct a_n's, hence it is related to $G(x)$. Actually, it equals $G(x)$ apart from the fact that it doesn't contain the term proportional to a_1: $G^2(x)$ begins with x^2. We can, therefore, introduce this 'initial condition' and write the recurrence relation eq.2.5 in terms of the generating function as

$$G(x) = a_1 x + \frac{1}{2}G^2(x) \tag{2.8}$$

This algebraic relation can be solved, $G(x)$ can be found in a closed form and then a Taylor expansion will yield all needed a_n. Alternatively, it can be seen as a recursive equation: inserting $G(x)$ at the right hand side provides an improved $G(x)$ at the left hand side, beginning with $G(x) = a_1 x$. This can be easily automated.

If the configuration D_n is formed by dividing the n 'things' in k groups each of which forms a D_{n_k}, one would immediately construct a recurrence relation using $G^k(x)$ (with a factor of $\frac{1}{k!}$ for permutations).

In general, whenever known recurrence relations exist in such problems, the method of generating functions is always preferable thanks to the way it automatically deals with the underlying combinatorics. Moreover, its merits extend beyond this book-keeping advantage since knowledge of the generating function in closed form can yield information about the asymptotic behavior of a_n, which, otherwise, has to be guessed from extrapolating the, often few, known a_n's . In the following sections we will make extensive use of this technique to deal with the counting of Feynman diagrams. In that case a_n is the number of Feynman diagrams with $n+1$ legs, in the amplitude $A_{a\to n}$.

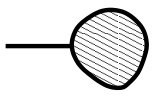
The Lagrangian of the theory, then, defines the type of allowed vertices in Feynman graphs. A ϕ^3 vertex, for example, would mean that an external line meets the rest of the diagram at a three-vertex, at which it 'splits in two'. This is equivalent to dividing the n remaining legs in two subgroups. Each subgroup with m external legs contains all sub-diagrams with $1 \to m$ legs, hence it is itself a legitimate subamplitude $A_{1\to m}$, for the purposes of counting[2]. We can, then, immediately construct a recurrence relation at the level of the generating function, involving its square. This relation is actually equivalent with the Schwinger-Dyson equation if the latter is stripped from all space-time or spin related information.

2.2.2 Counting tree diagrams

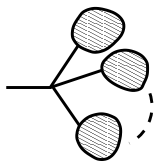
The counting of tree diagrams has been described in detail in [25, 29, 26], and here we briefly recapitulate these results. Let us denote by $a(n)$ the *number of tree-level Feynman diagrams* contributing to the $1 \to n$ amplitude, and define the generating function

$$\phi_0(x) = \sum_{n\geq 1} \frac{x^n}{n!}\, a(n) \quad . \tag{2.9}$$

Pictorially, we denote this as

$\phi_0(x) =$  (2.10)

By considering the alternatives when entering the blob from the left, we easily see that

$\phi_0(x) = \sum_{k\geq 2} \epsilon_{k+1}$  , (2.11)

[2] i.e. when the off-shellness of the external leg is of no interest.

where the right-hand side contains k blobs. This implies that $\phi_0(x)$ obeys the equation

$$\phi_0(x) = x + V'(\phi_0(x)) \quad . \tag{2.12}$$

Since $V(\varphi)$ is of order $\mathcal{O}(\varphi^3)$, this SD equation can easily be iterated starting with $\phi_0(x) = 0$, and the desired $a(n)$ can be read off once the iteration has proceeded far enough. Notice that

$$V''(\phi_0(x)) = 1 - \frac{1}{\phi_0'(x)} \quad , \quad V^{(p)}(\phi_0(x)) = \frac{1}{\phi_0'(x)} \frac{d}{dx} V^{(p-1)}(\phi_0(x)) \ \ (p \geq 3) \ \ , \tag{2.13}$$

so that the higher derivatives of $V(\phi_0(x))$ are completely expressed in terms of $\phi_0'(x)$ and *its* derivatives.

The asymptotic behavior of $a(n)$ for large n is determined by the singularity structure of $\phi_0(x)$. Since $\phi_0(x)$ cannot reach infinity for finite values of x, the singularities take the form of branch cuts, where $\phi_0(x)$ remains continuous but (as it turns out in all cases studied so far) $\phi_0'(x)$ diverges. We have

$$x = \phi_0 - V'(\phi_0) \quad \Rightarrow \quad \frac{dx}{d\phi_0} = 1 - V''(\phi_0) \ \ , \tag{2.14}$$

and the dominant singularity is reached for that value ϕ_c for which $V''(\phi_c) = 1$ and

$$x_c = \phi_c - V'(\phi_c) \tag{2.15}$$

is closest to the origin[3]. This value is always located on the positive real axis, where $\phi_0(x)$ is concave and monastically increasing for $x < x_c$. Taylor expansion then gives the structure of the branch cut:

$$\phi_0(x) \sim \phi_c - \sqrt{\frac{2x_c}{V^{(3)}(\phi_c)}} \left(1 - \frac{x}{x_c}\right)^{1/2} \quad , \tag{2.16}$$

from which we conclude that, for large n,

$$a(n) \sim \sqrt{\frac{x_c}{2\pi V^{(3)}(\phi_c)}} \ \frac{n!}{n^{3/2} x_c^n} = \frac{C}{\sqrt{4\pi}} \ \frac{n!}{n^{3/2} x_c^{n-1/2}} \quad . \tag{2.17}$$

with $C \equiv \sqrt{2/V^{(3)}(\phi_c)}$. In the table we give the relevant numbers for the four case theories.

theory	ϕ_c	x_c	C
φ^3	1	$1/2$	$\sqrt{2}$
φ^4	$\sqrt{2}$	$\sqrt{8/9}$	$2^{1/4}$
gQCD	$-1+\sqrt{3}$	$\sqrt{3} - 4/3$	$(4/3)^{1/4}$
effective	$\log(2)$	$2\log(2) - 1$	1

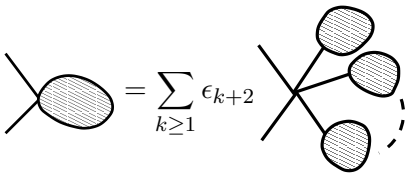
2.2.3 Counting one-loop diagrams

When closed loops are introduced, an SD-type equation itself cannot be used to count the number of *topologically inequivalent* graphs. This stems from the fact that the SD-type equations are *local* in the sense that they only consider (in a recursive manner) what happens at a single vertex of a diagram, while the topology of a graph containing closed loops is a *global* property of the whole graph. Instead, one has to settle for an order-by-order and topology-by-topology

[3]Here, we disregard the possibility that there are several such values, arising from a symmetry of the potential as in the case of theories with only φ^m interactions ($m \geq 4$). These cases are treated in detail in [28] and references therein. The asymptotic results given here are 'coarse-grained'.

treatment.

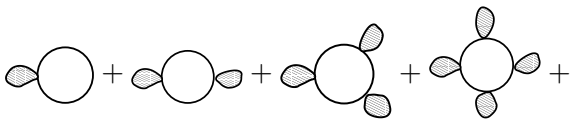
Every one-loop diagram can be viewed as a single closed loop, to which tree-diagram pieces (which we call *leaves*) are attached. From



$$= \sum_{k\geq 1} \epsilon_{k+2} = \sum_{k\geq 1} \frac{\epsilon_{k+2}}{k!}\phi_0(x)^k = V''(\phi_0(x)) \equiv v \ , \tag{2.18}$$

where the sum has k blobs again, and we have introduced the shorthand notation v, we see immediately that the number of one-loop graphs can be completely expressed in terms of v. The generating function of $L_1(n)$, the number of all one-loop non-vacuum graphs with precisely n external legs, is given by attaching leaves to a closed loop in all possible ways:

$$\begin{aligned} \mathcal{L}_1(x) &= \sum_{n\geq 1} \frac{x^n}{n!} L_1(n) \\ &= \end{aligned}$$



$$+ \cdots \tag{2.19}$$

The standard combinatorics for collecting the various external legs into leaves, and inspecting the symmetry properties of the resulting graphs, show that a one-loop graph with m leaves has precisely the 'natural' symmetry factor $1/(2m)$, with two important exceptions: the graphs with one or two leaves have an additional symmetry since, for the one-leave graph, the loop line may be flipped over, and for the two-leave graph the two internal loop lines may be interchanged. This leads us to the strategy for computing the number of topologically inequivalent graphs:

- Write down the vacuum graphs, with their 'natural symmetry factor';
- Attach leaves in all possible places;
- Multiply by the order of the residual symmetry left over after the particular attachment.

Performing this program for the one-loop case, we find

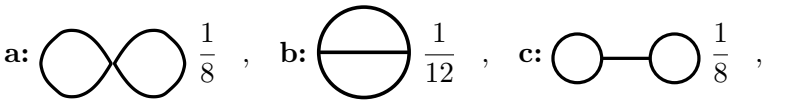
$$\begin{aligned} \mathcal{L}_1(x) &= \frac{2}{2}v + \frac{2}{4}v^2 + \sum_{m\geq 3} \frac{1}{2m} v^m = \frac{1}{2}v + \frac{1}{4}v^2 - \frac{1}{2}\log(1-v) \\ &= \frac{1}{2}v + \frac{1}{4}v^2 + \frac{1}{2}\log(\phi_0'(x)) \ . \end{aligned} \tag{2.20}$$

The number of one-loop diagrams with n external legs is given below for some theories.

N	φ^3	φ^4	gQCD	effective
1	1	0	1	1
2	2	1	3	3
3	7	0	14	15
4	39	7	99	111
5	297	0	947	1,104
6	2,865	145	11,460	13,836
7	33,435	0	167,660	209,340
8	457,695	6475	2,876,580	3,711,672
9	7,187,985	0	56,616,665	75,461,808
10	127,356,705	503,440	1,257,154,920	1,730,420,592

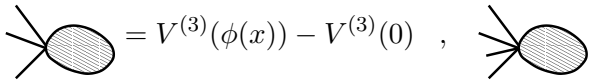
2.2.4 Counting two-loop diagrams

At the two-loop level, there are three topologically different vacuum diagrams. These are:



(2.21)

where we have indicated their 'natural' symmetry factor. Since these graphs contain vertices, we must also accommodate leaves attaching themselves to vertices:



(2.22)

In case no leave happens to be attached, the expression for the vertices read, of course, $V^{(3)}(0)$ and $V^{(4)}(0)$, respectively. This prohibits, for instance, the occurrence of a three-point vertex in a φ^4 theory. For each of the graphs we have to admit the possibility of zero, one, or more leaves on each line, and that of leaves on any vertex. For the determination of the residual symmetries it must be remembered that lines without leaves on them *may* be interchanged, and vertices without leaves *may* be interchanged, provided the 'anchoring' of the graph to the external legs contained in every leaf present permits such an interchange. As a simple example, the vacuum graphs themselves, without any leaves on them, have a residual symmetry of precisely 8,12, and 8, respectively, so that indeed they will be counted precisely one time. For graph **a** there are now $2 \times 3^2 = 18$ cases to be considered, and for **b** and **c** we have $2^2 \times 3^3 = 108$ cases. The results for their generating functions are:

$$\begin{aligned}
\mathcal{L}_2^{(\mathbf{a})}(x) &= \frac{1}{8} V^{(4)}(\phi) \left[\left(1 + v + \frac{1}{1-v} \right)^2 + 4 \right] , \\
\mathcal{L}_2^{(\mathbf{b})}(x) &= \frac{1}{12} \left[\left(V^{(3)}(\phi) \right)^2 \left(2 + \frac{3}{1-v} + \frac{1}{(1-v)^3} \right) \right. \\
&\qquad \left. + \left(V^{(3)}(0) \right)^2 \left(2 + 3(1+v) + (1+v)^3 \right) \right] , \\
\mathcal{L}_2^{(\mathbf{c})}(x) &= \frac{1}{8} \left[\left(V^{(3)}(\phi) \right)^2 \frac{1}{1-v} \left(1 + v + \frac{1}{1-v} \right)^2 \right. \\
&\qquad \left. + 4 \left(V^{(3)}(0) \right)^2 (1+v) \right] .
\end{aligned} \tag{2.23}$$

The total number $L_2(n)$ of two-loop graphs with precisely n external lines is therefore given via

$$\mathcal{L}_2(x) = \sum_{n \geq 0} \frac{x^n}{n!} L_2(n) = \mathcal{L}_2^{(\mathbf{a})}(x) + \mathcal{L}_2^{(\mathbf{b})}(x) + \mathcal{L}_2^{(\mathbf{c})}(x) \ . \tag{2.24}$$

Below, we give again the results for our specific theories.

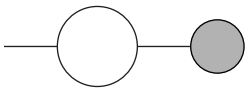
N	φ^3	φ^4	gQCD	effective
0	2	1	3	3
1	3	0	6	7
2	10	3	29	35
3	58	0	217	273
4	465	42	2,214	2,876
5	4,725	0	28,365	38,034
6	57,900	1,485	436,780	604,320
7	829,080	0	7,847,420	11,202,156
8	13,570,515	97,335	161,048,720	237,187,552
9	249,789,015	0	3,715,400,500	5,645,523,408
10	5,105,239,650	10,210,200	95,156,789,700	149,180,360,320

The extension to three or more loops is a matter of establishing the vacuum diagrams. For the three-loop case, however, there are 15 such graphs. Dressing them with leaves leads to a larger number of cases to be considered, ranging from 54 to 11,664 per graph.

2.2.5 Counting amputated diagrams

Loop diagrams containing tadpoles or seagulls are constant contributions to lower-order diagrams and are usually ignored. Moreover, diagrams containing self-energy loops on external legs are absorbed during the renormalization procedure. Removing such diagrams from the above results is a simple task. One has to subtract all contributions from (a) diagrams with loops carrying zero or one vertex and (b) diagrams carrying two vertices one of which is connected with an external leg while the other is a single propagator.

For the one-loop case one has to subtract the first graph in eq.(2.19), as a set of tadpole or seagull diagrams, as well as a contribution from graphs of the form



With these modifications the generating function reads

$$\mathcal{L}_1(x) = -\frac{1}{2}v + \frac{1}{4}v^2 - \frac{1}{2}\log(1-v) - x\phi_0(x)V^{(3)}(0) \tag{2.25}$$

The number of amputated one-loop diagrams for our test theories is given below.

N	φ^3	φ^4	gQCD	effective
3	1	0	4	4
4	12	3	39	43
5	117	0	437	502
6	1,290	75	5,800	6,916
7	16,425	0	90,450	111,660
8	239,400	3675	1,627,640	2,077,944
9	3,944,745	0	33,258,715	43,883,696
10	72,627,030	303240	761,405,820	1,037,955,824

For two-loops diagrams one has to consider each vacuum graph separately. All graphs containing loops with less than two vertices should be removed, as well as a variety of special cases which lead to non-amputated diagrams.

The generating functions for each of the three basic topologies becomes.

$$\begin{aligned}
\mathcal{L}_2^{\mathbf{(a)}}(x) &= \frac{1}{8}V^{(4)}(\phi)\left(\frac{v^2(2-v)^2}{(1-v)^2}\right) - V^{(4)}(0)V^{(3)}(0)x\phi_0(x) \\
\mathcal{L}_2^{\mathbf{(b)}}(x) &= \frac{1}{12}\left[\left(V^{(3)}(\phi_0)\right)^2\left(\frac{6-12v+9v^2-2v^3}{(1-v)^3}\right)\right] \\
&\quad -\frac{1}{12}\left(V^{(3)}(0)\right)^2\left[12x\phi_0 + \frac{6-3v^2+2v^3+v^4}{1-v}\right] \\
&\quad -\left(V^{(4)}(0)\right)^2 x\phi_0(x) - V^{(3)}(0)V^{(3)}(\phi_0) - V^{(3)}(0)V^{(4)}(0)2x\phi_0(x) \ ,
\end{aligned}$$

$$\mathcal{L}_2^{(\mathbf{c})}(x) = \frac{1}{8}\left[\left(V^{(3)}(\phi)\right)^2 \frac{v^2(2-v)^2}{(1-v)^3}\right] - \frac{1}{8}V^{(3)}(\phi)V^{(3)}(0)x\frac{4v(2-v)}{(1-v)^2} + \left(V^{(3)}(0)\right)^2 \frac{4}{8}\frac{x^2}{1-v} \tag{2.26}$$

The exact number of two-loop connected amputated diagrams for our test theories is given below.

N	φ^3	φ^4	gQCD	effective
3	4	0	28	37
4	63	9	457	600
5	870	0	7,285	9,760
6	12,945	460	128,675	177,160
7	212,940	0	2,552,165	3,617,824
8	3,874,815	35,315	56,538,055	82,588,784
9	77,605,290	0	1,387,411,690	2,089,438,256
10	1,700,078,625	4,090,800	37,407,699,175	58,096,995,744

2.2.6 Counting 1PI diagrams

The same methods as above can easily be employed in order to compute the number of one-particle irreducible (1PI) diagrams. We simply restrict ourselves to the 1PI vacuum bubbles; and, since 1PI diagrams cannot have any vertex in their leaves, we simply replace $\phi_0(x)$ in the arguments of $V'', V^{(3)}, V^{(4)}, \ldots$ by x. For the generating function of the 1PI one-loop diagrams, we therefore have

$$\mathcal{L}_1^{\mathrm{1PI}}(x) = \frac{1}{2}w + \frac{1}{4}w^2 - \frac{1}{2}\log(1-w) \quad , \quad w = V''(x) \ . \tag{2.27}$$

The resulting numbers are given in the following table.

N	φ^3	φ^4	gQCD	effective
1	1	0	1	1
2	1	1	2	2
3	1	0	4	5
4	3	3	12	17
5	12	0	57	83
6	60	15	390	557
7	360	0	3,195	4,715
8	2,520	315	30,555	47,357
9	20,160	0	333,900	545,963
10	181,440	11,340	4,105,080	7,087,517

At the two-loop level, we similarly find

$$\mathcal{L}_2^{\mathrm{1PI}}(x) = \frac{1}{8}V^{(4)}(x)\left[\left(1+w+\frac{1}{1-w}\right)^2+4\right] + \frac{1}{12}V^{(3)}(x)^2\left[2+\frac{3}{1-w}+\frac{1}{(1-w)^3}\right] + \frac{1}{12}V^{(3)}(0)^2\left[2+3(1+w)+(1+w)^3\right] \ . \tag{2.28}$$

Numbers are given below.

N	φ^3	φ^4	gQCD	effective
0	1	1	2	2
1	1	0	3	4
2	2	2	9	13
3	7	0	40	62
4	36	12	265	410
5	240	0	2,230	3,499
6	1,860	225	22,485	36,213
7	16,380	0	261,135	435,852
8	161,280	8,295	3,418,695	5,944,000
9	1,753,920	0	49,712,670	90,309,029
10	20,865,600	481,950	794,102,400	1,510,208,963

2.3 Counting with symmetry factors

The counting of diagrams including their symmetry factors is a somewhat simpler task, which can be performed on the basis of the path integral itself. In [54] this has been discussed in detail. However our approach here is somewhat different. One can expand the generating function of the number of connected diagrams perturbatively around $\varphi = 0$ and get a series in x (the source). Or, alternatively, one can expand perturbatively around the tree-level one-point function $\varphi = \phi_0$. This shift eliminates the source x in favor of the tree-level one-point function $\phi_0(x)$, and reveals the vacuum-graph dressing procedure that we employed above.

2.3.1 Counting diagrams with symmetry factors

Consider the generating function for the number of disconnected diagrams of a scalar theory with arbitrary couplings and a source x:

$$Z(x) = N \int d\varphi \;\; exp\left(-\frac{1}{\hbar}(\frac{1}{2}\varphi^2 - V(\varphi) + x\varphi)\right) \tag{2.29}$$

with $N = 1/\sqrt{2\pi\hbar}$. Expanding around the tree-level approximation ϕ_0 of the one-point function, i.e. setting $\varphi \to \phi_0 + \varphi$, and making use of the Schwinger -Dyson equation for $\phi_0(x)$ gives

$$Z(x) = N \;\; exp(-\frac{1}{\hbar}S(\phi_0) + \frac{x\phi_0}{\hbar}) \int d\varphi \; e^{-\frac{1}{\hbar}\hat{S}(\varphi)} \tag{2.30}$$

with

$$\hat{S}(\varphi) = \frac{1 - V''(\phi_0)}{2}\varphi^2 - \sum_{n=3}^{\infty} V^{(n)}(\phi_0)\frac{\varphi^n}{n!} \tag{2.31}$$

The generating function of the number of connected diagrams is then

$$W(x) = \hbar \log(Z(x)) = -S(\phi_0) + x\phi_0 + \hbar \log(N \int d\varphi e^{-\frac{1}{\hbar}\hat{S}(\varphi)}) \tag{2.32}$$

We see that it can be seen as a sum of the tree-level part plus higher order corrections. These corrections can be written as the generating function for the vacuum diagrams of a theory with action $\hat{S}(\varphi)$. The Feynman rules corresponding to this action can be read off directly :

- $\frac{1}{1-V''(\phi_0)} = \phi_0'$ for every propagator.
- $V^{(n)}(\phi_0)$ for every n-point vertex.

Given the potential $V(\varphi)$ of the theory one can expand the vertex terms in the exponential of eq.(2.32), calculate the Gaussian integrals and arrive at an expression for $W(x)$ that contains only $V''(\phi_0(x))$ and its derivatives. In this way, given the tree-level one-point function of the theory, one finds the number of graphs weighted by their symmetry factors to arbitrary order.

Writing $W(x)$ in an $\hbar$ expansion

$$W(x) = W_0(x) + \hbar W_1(x) + \hbar^2 W_2(x) + \dots \tag{2.33}$$

and, performing the integral and collecting together the terms of the same order in $\hbar$, we see that the one loop diagrams are generated by

$$W_1(x) = \frac{1}{2}\log(\frac{1}{1 - V''(\phi_0)}) \tag{2.34}$$

We can also find the generating function for the two-loop diagrams

$$W_2(x) = \frac{1}{8}\frac{V^{(4)}(\phi_0)}{(1 - V''(\phi_0))^2} + \frac{5}{24}\frac{V^{(3)}(\phi_0)V^{(3)}(\phi_0)}{(1 - V''(\phi_0))^3} \tag{2.35}$$

The factor $\frac{1}{8}$ in front of the first term is the symmetry factor of the only 2-loop vacuum diagram with a four-vertex (see fig 2.21.a). The factor $\frac{5}{24} = \frac{1}{8} + \frac{1}{12}$ is the sum of the symmetry factors of the two vacuum diagrams with two three-vertices (see fig 2.21 .b and .c)[4].

Writing the derivatives $V^{(m)}(\phi_0)$ in terms of derivatives of ϕ_0 (which can be done by differentiating the Schwinger-Dyson equation for ϕ_0) one arrives at

$$W_2(x) = \frac{1}{8}\frac{\phi_0'''}{(\phi_0')^2} - \frac{1}{6}\frac{(\phi_0'')^2}{(\phi_0')^3} \tag{2.36}$$

Below, we give results for our four case theories in one loop.

N	φ^3	φ^4	gQCD	effective
1	1/2	0	1/2	1/2
2	1	1/2	3/2	3/2
3	4	0	15/2	8
4	24	7/2	57	63
5	192	0	1,149/2	658
6	1,920	80	7,230	8,568
7	23,040	0	218,175/2	133,676
8	322,560	3,815	1,919,190	2,430,816
9	5,160,960	0	77,146,125/2	50,484,016
10	92,897,280	31,0940	871,927,770	1,178,963,856

The results for the four theories in two loops are again collected below.

[4]In fact one could even avoid performing the integral since the generating function for N loops is simply the sum of the vacuum graphs with N loops weighted by their symmetry factors using the Feynman rules for the $\hat{S}(\varphi)$ action given above. However, this presupposes that one knows what the symmetry factor of the specific vacuum diagram is.

N	φ^3	φ^4	gQCD	effective
1	5/8	0	31/24	17/12
2	25/8	2/3	25/3	19/2
3	175/8	0	1,777/24	527/6
4	1,575/8	149/12	5,057/6	1,037
5	17,325/8	0	280,735/24	44,726/3
6	225,225/8	1,535/3	1,149,515/6	252,734
7	3,378,375/8	0	86,813,545/24	14,808,232/3
8	57,432,375/8	111,755/3	464,096,885/6	109,143,424
9	1,091,215,125/8	0	44,344,732,495/24	8,085,390,392/3
10	22,915,517,625/8	12,672,800/3	292,590,237,275/6	73,514,104,288

Both in the one and two-loop cases an intriguing pattern of denominators is apparent for large N values, which seems to persist (we have checked this for N up to 50).

The above procedure can easily be extended to higher-loop amplitudes as well, but since we have not computed the unweighted diagram sums we defer this discussion to the case of asymptotically large N.

2.3.2 counting 1PI graphs

The generating function for the one particle irreducible diagrams of a theory weighted by their symmetry factors can be obtained by the same prescription by substituting $\phi_0 = x$. Now, however, we have to take into account only the 1PI vacuum diagrams. In the one loop case the only vacuum graph is 1PI and the generating function is

$$W_1(x) = \frac{1}{2}\log(\frac{1}{1-V''(x)}) \tag{2.37}$$

In the two loop case we have to take into account the vacuum graph with one 4-vertex (see figure 2.21.a) and only one of the two vacuum graphs with three vertices (see figure 2.21.b) since the other one (see figure 2.21.c) is not 1PI . This alters the symmetry factor from $\frac{5}{24}$ to $\frac{1}{12}$. We get then

$$W_2(x) = \frac{1}{8}\frac{V^{(4)}(x)}{(1-V''(x))^2} + \frac{1}{12}\frac{V^{(3)}(x)V^{(3)}(x)}{(1-V''(x))^3} \tag{2.38}$$

We give below the number of irreducible diagrams weighted by their symmetry factors in the 1-loop case for the four test theories :

	φ^3	φ^4	$\varphi^3+\varphi^4$	effective
N=1	1/2	0	1/2	1/2
N=2	1/2	1/2	1	1
N=3	1	0	5/2	3
N=4	3	3/2	21/2	13
N=5	12	0	57	75
N=6	60	15	390	541
N=7	360	0	3,195	4,683
N=8	2,520	315	30,555	47,293
N=9	20,160	0	333,900	545,853
N=10	181,440	11,440	4,105,080	708,7261

We give below the number of irreducible diagrams weighted by their symmetry factors in the 2-loop case for the four test theories :

	φ^3	φ^4	$\varphi^3+\varphi^4$	effective
N=1	1/4	0	2/3	19/24
N=2	1	5/12	41/12	101/24
N=3	5	0	89/4	691/24
N=4	30	21/4	709/4	5765/24
N=5	210	0	1,660	56,659/24
N=6	1,680	135	17,865	64,0421/24
N=7	15,120	0	217,035	8,178,931/24
N=8	151,200	5,775	2,936,745	116,422,085/24
N=9	1,663,200	0	43,787,520	1,827,127,699/24
N=10	19,958,400	368,550	713,163,150	31,336,832,741/24

2.4 Asymptotic estimates

It is fairly easy to estimate the number of diagrams, both with and without their symmetry factors, for asymptotically large N. As before, the asymptotic behavior of these numbers is governed by the analytic structure of their generating functions close to that singularity which is closest to the origin (that is, around $x \sim x_c$). There, we have

$$\phi_0'(x) \sim \frac{1}{2} C \, (x_c - x)^{-1/2} \quad , \quad C = \left(2/V^{(3)}(\phi_c)\right)^{1/2} \quad , \tag{2.39}$$

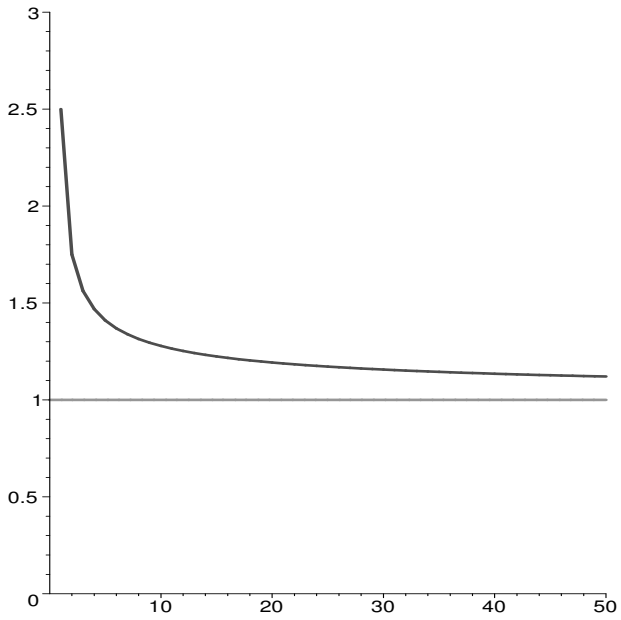
where x_c, ϕ_c and C again depend on the theory. Let us first concentrate on the one-loop diagrams. Since $v = 1 - 1/\phi_0'(x)$ has a square-root branch cut at the singular point, $\log(1-v)$ is more singular than v or v^2, and we have

$$\mathcal{L}_1(x) \sim \frac{1}{2} \log\left(\frac{C}{2\sqrt{x_c - x}}\right) = \mathcal{L}_1^{(s)}(x) \ . \tag{2.40}$$

We conclude that, for one-loop diagrams, the *average symmetry factor* of a given diagram is asymptotically equal to 1. The number $K_1(N)$ of graphs contained in the one-loop N-point amplitude is asymptotically given by

$$K_1(N) \sim \frac{1}{4} \frac{1}{(x_c)^N} \frac{N!}{N} \tag{2.41}$$

To illustrate the convergence of the weighted number of graphs to the unweighted number, we give the ratio of the coefficients of x^N in $\mathcal{L}_1^{(s)}(x)$ to those of $\mathcal{L}_1(x)$ as a function of N below, for the pure φ^3 theory. The other cases show a similar behavior[5], in which the asymptotic regime is approached as $1/\sqrt{N}$: this can also be easily checked from the exact form of $\mathcal{L}_1(x)$ close to the singularity.



The asymptotic results for the higher-loop amplitudes can be established by the following reasoning. The leading contribution from each leave-dressed vacuum diagram is given by that part that has the highest degree of divergence as $x \to x_c$. From each line in the vacuum graph, this is a factor $1/(1-v) = \phi_0'(x)$. Furthermore, from each k-point vertex in the vacuum graph the

[5] For the pure φ^4 theory, this holds in the 'coarse-grained' approximation [28].

leading contribution comes from the limiting behavior of $V^{(k)}(\phi_0(x))$. Now, it is easily seen that, as $x \to x_c$,

$$V^{(3)}(\phi_0(x)) \sim \frac{2}{C^2} \quad \Rightarrow V^{(k)}(\phi_0(x)) \sim 0 \ , \ \ k \geq 4 \ . \tag{2.42}$$

We conclude that the leading behavior of the number of unweighted graphs is given by those vacuum graphs that contain only three-point vertices. To get the number of unweighted diagrams at the L-loop level, therefore, we first compute the normalized path integral for the pure φ^3 theory, using the usual perturbative interchange between expansion of the potential term and integration:

$$\begin{aligned} Z &= N \int\limits_{-\infty}^{\infty} d\varphi \, \exp\left(-\frac{\mu}{2}\varphi^2 + \frac{\lambda}{6}\varphi^3\right) \\ &= \sum_{n\geq 0} \frac{(6n)!}{(2n)!(3n)!(576)^n} \left(\frac{\lambda^2}{\mu^3}\right)^n \ . \end{aligned} \tag{2.43}$$

The sum of all connected vacuum diagrams with interactions is then given by

$$W = \log(Z) \ , \tag{2.44}$$

in the expansion of which the L-loop contribution ($L \geq 2$) is given by the term with λ^{2L-2}. In this expression, it suffices to replace λ by $2/C^2$ and μ by $1/\phi_0'(x)$. The result is

$$W = \sum_{L\geq 2} w_L \, C^{1-L} \, (x_c - x)^{3(1-L)/2} \ . \tag{2.45}$$

The first coefficients w_L are given below.

L	w_L	L	w_L
2	5 / 48	7	19675 / 6144
3	5 / 64	8	1282031525 / 88080384
4	1105 / 9216	9	80727925 / 1048576
5	565 / 2048	10	1683480621875 / 3623878656
6	82825 / 98304	11	13209845125 / 4194304

The asymptotic result for $K_L(N)$, the number of unweighted diagrams contributing to the L-loop n-point amplitude is therefore given by

$$K_L(N) \sim \frac{\Gamma\left(N + \frac{3}{2}(L-1)\right)}{\left(x_c^{3/2} C\right)^{L-1} \Gamma\left(\frac{3}{2}(L-1)x_c^N\right)} \ . \tag{2.46}$$

For the number of L-loop graphs weighted by their symmetry factors we may employ the following formulation of the SD equation:

$$\phi_L = \sum_{\{n_{p,q}\}\geq 0} V^{(m)} \prod_{p,q\geq 0} \frac{1}{(n_{p,q})!} \left(\frac{1}{(q+1)!}\phi_p^{(q)}\right)^{n_{p,q}} \ , \tag{2.47}$$

where the bracketed superscripts denote derivatives, and

$$\sum_{p,q} (p+q) n_{p,q} = L \quad , \quad m = 1 + \sum_{p,q} (q+1) n_{p,q} \ . \tag{2.48}$$

The successive expressions for $\phi_L(x)$ in terms of lower-loop ones can straightforwardly be worked out. For $L = 1, 2$ these have been given in the previous section. If we now put in the approximate

form of $\phi_0(x)$ given in eq.(2.16), it is easily checked (at least up to $L = 10$) that expression W is reproduced. Note that in this approximation the fourth and higher derivatives of $V(\phi_0)$ vanish, so that Eq.(2.47) is actually more complicated than need be: nevertheless, by using the next-to-leading expression

$$\phi_0(x) \sim \phi_c - C(x_c - x)^{1/2} - C'(x_c - x) \ , \tag{2.49}$$

it can also be checked that, indeed, the subleading behavior of $\phi_0(x)$ shows up only in the subleading terms in $K_L(N)$. We conclude that *as $N \to \infty$, the average symmetry factor of any Feynman diagram approaches unity.*

2.5 Complexity of the Caravaglios-Moretti algorithm

2.5.1 The Caravaglios-Moretti algorithm

The Caravaglios-Moretti (CM) algorithm, as first explicitly given in [16] (and earlier implied by [6]), is an algorithm for recursive computations of tree-level amplitudes based on the Schwinger-Dyson equation[6], eq. 1.23. In any given calculation, on a diagram-per-diagram base, one has to compute repeatedly identical pieces of Feynman graphs. The main concept in the CM approach is to compute such off-shell subamplitudes once, and use them as building blocks for bigger subamplitudes, until the whole amplitude is built.

More precisely, an arbitrary external leg is selected as a starting point for the algorithm. Let us call it particle 1. As a next step, assuming that the theory includes a three-vertex interaction, all the possible subamplitudes where a particle $i \neq 1$ is connected with any other particle $j \neq 1$ with a three-vertex, are computed. These subamplitudes will, now, have an off-shell leg, carrying the sum of the momenta of the two particles from which they are formed. In a scalar theory this subamplitude is a scalar function. In theories with fermions or spin-1 bosons, the subamplitude can also be a spinor function ($\Psi^\alpha(p_i, p_j)$ or $\overline{\Psi}^\alpha(p_i, p_j)$) or a four-vector function ($M^\mu(p_i, p_j)$) and they can also carry other internal degrees of freedom (like color).

Next, all subamplitudes containing two external legs (also called level 2 subamplitudes) are combined in all possible ways with one of the remaining external legs to form subamplitudes of level three, etc. until the unique subamplitude of level $N-1$, containing all legs except 1, is reached. This has to be contracted with the external leg of particle 1 to form the amplitude.

In case a four-vertex is also present, subamplitudes of level three, for example, can also be formed by combining three external legs. This increases the computational complexity of the algorithm, so a work-around has been devised for QCD-like theories that uses an auxiliary field to break a four-vertex into three-vertices, see [28] for details.

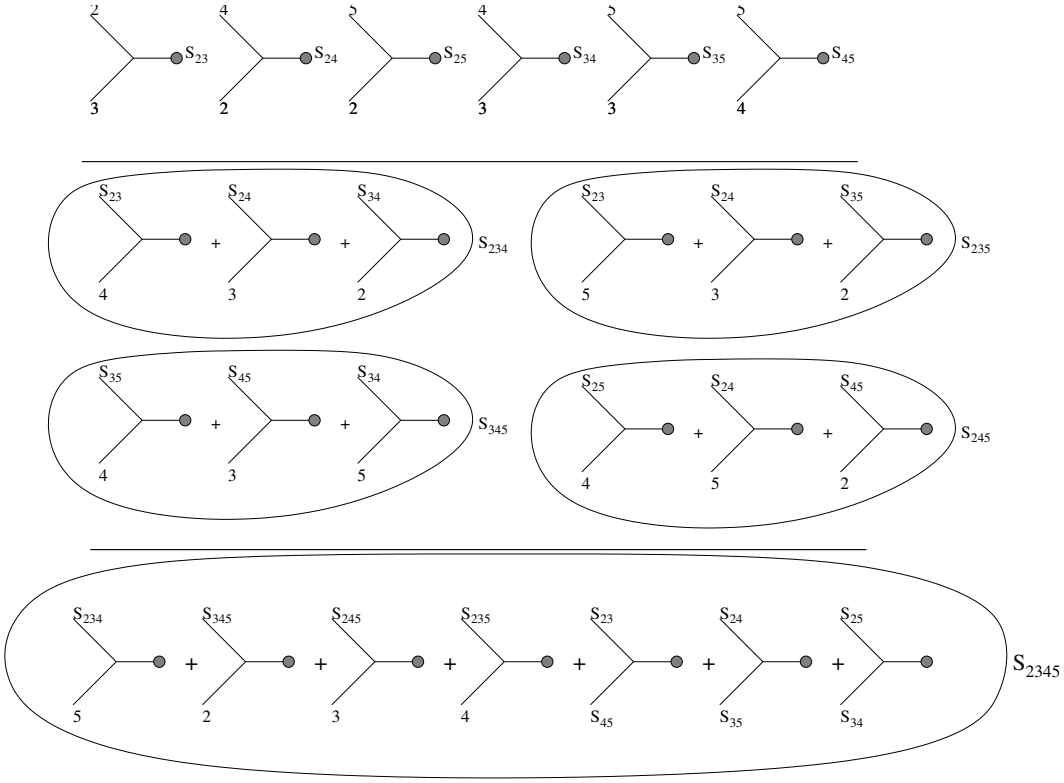
A diagrammatic example of the first five recursion steps is given in Table 2.1, for a scalar theory with three-vertices. We have denoted with S_n the subamplitudes of level n and the presence of an external leg by p_i. For more detailed descriptions, and examples involving QCD we refer to [16, 28, 55]: in what follows we are only interested in the combinatorics of the algorithm.

2.5.2 Complexity for tree-level computations in any theory

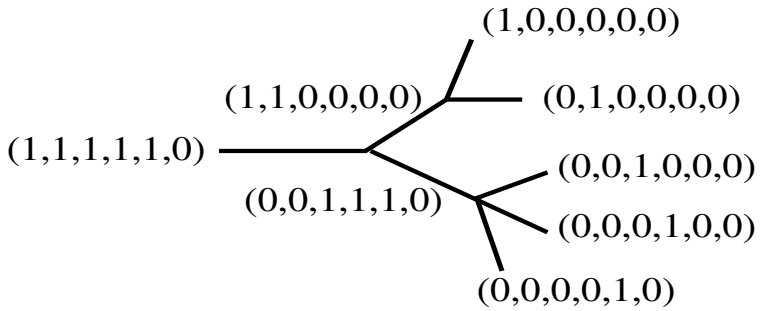
We assume an N-particle process, and set $K = N-1$. Each subamplitude can then be encoded by a binary string with N bits, each referring to a given external particle. The bit is set to 1 if its external leg is involved in the subamplitude, and to 0 otherwise. For instance, the string $(1,1,0,1,1,0,0,0,\ldots,0,0)$ denotes that subamplitude in which the external particles with labels 1,2,4 and 5 are combined, using the vertices of the theory, into a single off-shell momentum. By the same convention, a string with a single 1 refers to the Feynman rule for a single external particle (a spinor or antispinor for fermions, a polarization vector for vector particles, etc.). The CM algorithm combines subamplitudes into successively more complicated ones, culminating in the string $(1,1,1,\ldots,1,1,1,0)$, which, after multiplying with the external factor $(0,0,0,\ldots,0,0,0,1)$ gives the final answer for the amplitude. It is clear that of the N external particles, one can be left out of the combinatorics since it has to be included only at the very end. The combinatorial problem is, therefore, to determine the number of ways to

[6] In the original paper, [16], the recursion relations are derived from the effective action. The observation that these relations are closely related to the Schwinger-Dyson equation was first made in [39].

Table 2.1: The levels of recursion of the CM algorithm for a ϕ^3 theory.



decompose a string of K bits. An example of a possible decomposition for $K = 5$ is



where we have indicated the strings corresponding to the external legs and the various subamplitudes. The possible decompositions depend on the theory in question: the presence of an $(m + 1)$-point vertex in the theory allows for a decomposition into m smaller strings. In this chapter, we shall only deal with theories of a single self-interacting field (gluonic QCD being a particularly interesting example): extensions to more fields are fairly straightforward. In recent implementations such as **HELAC** [39], this decomposition can be recognized explicitly.

Let us first consider a subamplitude's string with n 1's being decomposed into m smaller strings, each with at least one 1. This happens when, in the SD equation, an $(m + 1)$-point vertex is encountered. The number of inequivalent decompositions of a subamplitude of level n in m pieces, denoted by $c_m(n)$, is given by

$$c_m(n) = \frac{1}{m!} \sum_{n_{1,2,\ldots,m} \geq 1} \frac{n!}{n_1!\, n_2!\, \cdots\, n_m!} \ , \tag{2.50}$$

where, of course, $n_1 + n_2 + \cdots + n_m = n$. Note that the above equation assumes that all the subamplitudes containing $n_1, n_2, \ldots, n_m$ external momenta exist. This is always the case when a φ^3 interaction is present in the theory[7] but it is not true for a pure φ^4 theory for example. Then one has to introduce a factor that cancels the terms coming from combinations of non-allowed subamplitudes. We, nevertheless , proceed with our program to find a generating function for $c_m(n)$, having in mind effective theories that always contain a three point vertex. We find

$$\sum_{n\geq 0} \frac{x^n}{n!} c_m(n) = \frac{1}{m!} \left(e^x - 1\right)^m \tag{2.51}$$

Since the full amplitude consists of $N = K + 1$ external legs, we have K legs to combine, so the strings describing the subamplitudes have length K. Now, out of all bit strings of size K, there are precisely $K!/n!(K-n)!$ strings containing precisely n 1's (or, equivalently, there are $\left(\begin{array}{c} K \\ n \end{array} \right)$ subamplitudes of level n). The total number of decompositions in the algorithm (for an amplitude with $K+1$ legs) involving $(m+1)$-point vertices is therefore found by summing $\left(\begin{array}{c} K \\ n \end{array} \right) c_m(n)$ over all n (i.e. over all levels)

$$f_m(K) = \sum_{n\geq 0} \left(\begin{array}{c} K \\ n \end{array} \right) c_m(n) \ , \tag{2.52}$$

The generating function of $f_m(K)$ is

$$g_m(x) \equiv \sum_{K\geq 0} \frac{x^K}{K!} f_m(K) = \frac{1}{m!} e^x \left(e^x - 1\right)^m \ . \tag{2.53}$$

In the simple case of a pure φ^3 theory we therefore have

$$g_2(x) = \frac{1}{2}\left(e^{3x} - 2e^{2x} + e^x\right) = \sum_{K\geq 0} \frac{x^K}{K!} \frac{1}{2}\left(3^K - 2^{K+1} + 1\right) \ , \tag{2.54}$$

so that the number of decompositions necessary to arrive at an N-point amplitude is given by

$$\frac{1}{6}3^N - \frac{1}{2}2^N + \frac{1}{2} \ .$$

For a theory with both φ^3 and φ^4 interactions such as gluonic QCD, we find a total of

$$\frac{1}{24}4^N - \frac{1}{4}2^N + \frac{1}{3}$$

decompositions.

In QCD at the tree-level, an improvement is possible. We can decompose the gluonic four-vertex into two 3-vertices by employing an auxiliary field, as explained for instance in [28]. This brings the complexity down from 4^N to 3^N, a worthwhile improvement for large N. It is not to be expected, however, that this will be possible in higher orders. The effective action, therefore, will contain $(m+1)$-vertices for all $m \geq 2$, and the generating function is therefore

$$F(x) = \sum_{m\geq 2} g_m(x) = \exp\left(e^x - 1 + x\right) - \exp(2x) \ . \tag{2.55}$$

Below we give the number of decompositions,

$$D(N) = \sum_{m\geq 2} g_m(N-1) \ , \tag{2.56}$$

for not-too-large values of N.

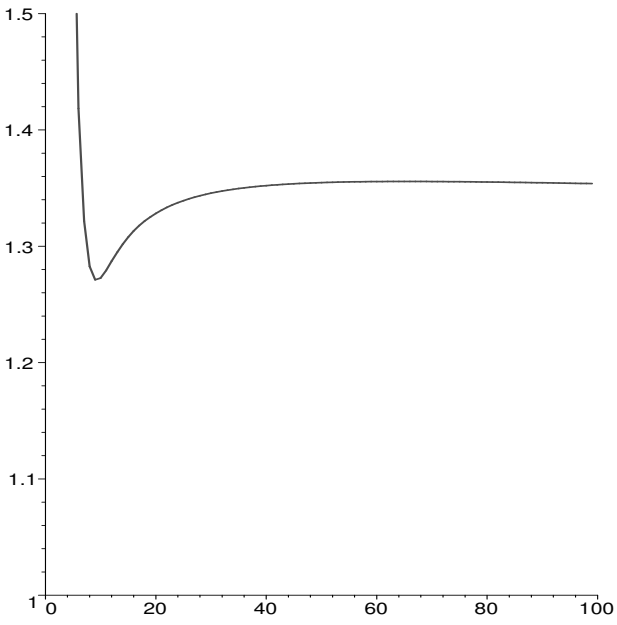
[7] Because then there is always the possibility of constructing a subamplitude containing n_k external momenta by combining a subamplitude containing $n_k - 1$ momenta with an external momentum in a three point vertex.

N	$D(N)$	N	$D(N)$
3	1	8	4,012
4	7	9	20,891
5	36	10	115,460
6	171	11	677,550
7	813	12	4,211,549

For asymptotically large values of N, we have to study the analytic structure of $F(x)$. Since this function is analytic for finite x, $D(N)$ must increase with N slower than $N!$. On the other hand, $D(N)$ increases faster than c^N for any finite c, which is reasonable since as N grows larger and larger values of m come into play. This is also evident from the fact that the standard Borel transform of the series $F(x)$,

$$\int_0^\infty dy\, e^{-y}\, F(xy) = -\frac{1}{1-2x} + \int_0^\infty dy\, \exp\left(-y + e^{xy} - 1 + xy\right) \tag{2.57}$$

does not converge for any positive value of x.



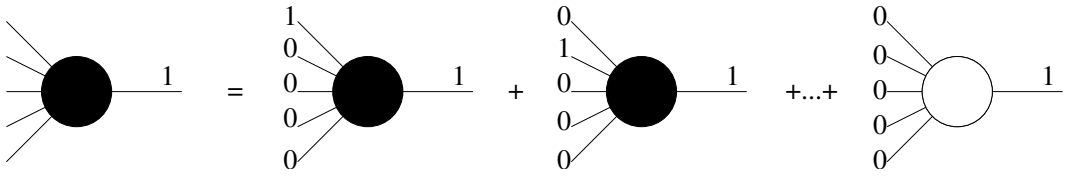
The above plot shows the behavior of the ratio $(\log(N)/N)(D(N)/D(N-1))$ as a function of N for $3 \leq N \leq 100$. For high N, this ratio is approximately (but not quite) a constant.

2.5.3 Complexity in one and two loops

Consider a general theory with m-point vertices. Each subamplitude of level n (containing n specific external momenta), can be constructed by combining two or more lower-level subamplitudes in a three- or more-point vertex. When using an $(m+1)$-point vertex the subamplitude is built by m lower level subamplitudes and the number of different ways for this to happen is given by eq.(2.50).

Each term in the series represents the number of ways to construct the subamplitude of level n using subamplitudes of level $n_1, \ldots, n_m$. The computational cost of any such subamplitude involves (assuming that there is an $m+1$ vertex in the theory) contributions from the following possibilities: all lower subamplitudes are free of loop corrections and the vertex is an ordinary

one (this gives the tree-level subamplitude)[8]. It can also be that one of the subamplitudes already contains a loop correction (that occurred in previous steps in the CM algorithm) and the vertex is an ordinary one (see fig 2.58). The subamplitudes containing loop corrections can, however, be of level two or higher since the level one subamplitudes are the external legs which we consider amputated. There are, therefore, $m - \sum_i \delta_{1,n_i}$ different possibilities. Finally there is the case that all subamplitudes are free of loop corrections but the vertex is actually a loop (see the last term in fig 2.58). The number of different possibilities is now equal to the number of 1PI diagrams with one loop and $m+1$ legs, which we denote with $J_{m,1}$.



(2.58)

The cost of computing the specific subamplitude via an $m+1$ vertex is therefore

$$\frac{1}{m!}\sum_{n_1,\ldots,n_m}\frac{n!}{n_1!n_2!\ldots n_m!}\delta(\sum n_i - n)(V_m(1+m-\sum_i \delta_{1,n_i}) + J_{m,1}) \tag{2.59}$$

where we have included a factor $V_m = 1$ if the $m+1$ vertex is in the theory and $V_m = 0$ if not. The cost of the subamplitude is then found by summing over m. There are $\frac{(N-1)!}{n!(N-1-n)!}$ different subamplitudes. The computational cost of the whole algorithm in units of effective vertices is then

$$\sum_{n=2}^{K}\left(\begin{array}{c} K \\ n \end{array}\right)\sum_{m\geq 2}^{\infty}\frac{1}{m!}\sum_{\substack{n_1,\ldots,n_m \\ \sum n_i = n}}\frac{n!}{n_1!\ldots n_m!}(V_m(1+m-\sum_i \delta_{1,n_i}) + J_m) \tag{2.60}$$

where $K = N-1$. In the following table we present the results for the four test theories.

	φ^3	φ^4	$\varphi^3+\varphi^4$	effective
N=1	0	0	0	0
N=2	0	0	0	0
N=3	2	0	5	6
N=4	18	4	46	57
N=5	114	0	340	442
N=6	720	105	2,715	3,713
N=7	5,368	0	26,346	37,411
N=8	49,686	3,395	315,035	459,056
N=9	553,766	0	4,474,868	6,688,320
N=10	7,112,700	149,140	72,741,355	112,139,709

In order to include the two-loop correction one has to add to the above formula a term $(m-\sum \delta_{1,n_i})(m-\sum \delta_{n_i}-1)$ for the possibility that two of the lower subamplitudes have a one-loop correction and a term equal to $m-\sum \delta_{1,n_i}$ for the possibility that one of the subamplitudes has a two-loop correction. There is also the possibility that one of the lower subamplitudes is

[8]That is provided that the lower-level subamplitudes exist! This always happens when the theory involves φ^3 interactions. In the pure φ^4 theory, however, we have to modify the calculation to exclude combinations where one of the n_i's is equal to 2 since in such a theory there are no level two subamplitudes.

of one-loop order and the vertex itself is a one-loop 1PI graph. This costs an extra term of $J_{m,1}(m - \sum \delta_{1,n_i})$ Moreover one has to add the number of 1PI graphs with 2 loops and $m+1$ legs, $J_{m,2}$. Hence we now have, writing $S_{n_i} = \sum_i \delta_{1,n_i}$

$$\sum_{n=2}^{K} \left(\begin{array}{c} K \\ n \end{array} \right) \sum_{m \geq 2}^{\infty} \frac{1}{m!} \sum_{\substack{n_1,\ldots,n_m \\ \sum n_i = n}} \frac{n!}{n_1! \ldots n_m!} (A + B + C)$$

where $K = N - 1$ and

$$\begin{aligned} A &= V_m(1 + m - S_{n_i} + (m - S_{n_i})(m - S_{n_i} - 1) + m - S_{n_i}) \\ B &= J_{m,1} + J_{m,1}(m - S_{n_i}) \\ C &= J_{m,2} \end{aligned}$$

The results for the four test cases are presented below.

	φ^3	φ^4	$\varphi^3 + \varphi^4$	effective
N=1	0	0	0	0
N=2	0	0	0	0
N=3	9	0	45	68
N=4	102	16	566	857
N=5	957	0	6,414	9,837
N=6	9,740	610	81,560	127,451
N=7	114,677	0	1,201,556	1,920,824
N=8	1,546,986	32,151	20,211,345	33,181,094
N=9	23,395,461	0	380,938,056	644,468,452
N=10	390,310,512	2,574,670	7,929,937,496	13,861,514,611

One should be aware of the fact that the above results are obtained under the assumption that the computational cost for every effective vertex that might include one- or two-loop 1PI graphs is the same.

2.6 Comparison of the complexity of the CM algorithm to the diagrammatic approach

We present below the ratio of the computational complexity of the CM algorithm over the number of diagrams one has to calculate in the customary diagrammatic approach, for our four test theories[9]. For each case the ratio for a calculation in tree, tree plus 1-loop, and tree plus 1- and 2- loop level is presented.

complexity of CM algorithm / number of diagrams						
	φ^3			φ^4		
	L_0	L_0+L_1	$L_0+L_1+L_2$	L_0	L_0+L_1	$L_0+L_1+L_2$
N=3	1.00	1.000	1.500	-	-	-
N=4	2.00	1.200	1.307	1.000	1.000	1.231
N=5	1.666	0.864	0.955	-	-	-
N=6	0.857	0.516	0.679	2.00	1.235	1.119
N=7	0.318	0.309	0.498	-	-	-
N=8	0.093	0.199	0.375	1.575	0.858	0.819
N=9	0.022	0.136	0.286	-	-	-
N=10	0.005	0.095	0.220	0.636	0.468	0.584

complexity of CM algorithm / number of diagrams						
	$\varphi^3+\varphi^4$			effective		
	L_0	L_0+L_1	$L_0+L_1+L_2$	L_0	L_0+L_1	$L_0+L_1+L_2$
N=3	1.000	1.000	1.363	1.000	1.200	1.619
N=4	1.500	1.070	1.132	1.500	1.212	1.325
N=5	1.00	0.736	0.828	0.961	0.837	0.956
N=6	0.409	0.451	0.605	0.381	0.519	0.691
N=7	0.121	0.283	0.454	0.109	0.327	0.515
N=8	0.028	0.190	0.347	0.025	0.217	0.392
N=9	0.005	0.132	0.268	0.005	0.150	0.302
N=10	0.001	0.094	0.208	0.001	0.107	0.234

One should note that the CM algorithm will actually perform better than depicted by the above numbers, when compared to the straightforward diagrammatic approach, since we consider the cost of a step in the CM algorithm (i.e. the calculation of a subamplitude which corresponds to the calculation of an effective vertex) equal to the cost of the computation of a whole diagram. That is the reason for the apparently poor performance of the CM algorithm in the case of tree level φ^4 theory.

2.7 Conclusions and outlook

We have seen that the Caravaglios-Moretti algorithm is more efficient than the straightforward diagrammatic approach, in the tree as well as the one and two loop level, by a factor that increases rapidly with the number of external legs, even though this increase is less rapid in the one- and two- loop level than in tree level. This conclusion is not expected to change when fermionic degrees of freedom are included in the calculation. It seems, therefore, that the CM algorithm is at present the preferred method of performing multi-leg tree-level as well as multi-loop calculations, even if no better approach than the straightforward effective vertex one assumed here is ever discovered.

[9] Only amputated, tadpole/seagull-free diagrams are considered

The CM algorithm, along with its variant for color-stripped amplitudes [7], can easily and efficiently be implemented in computer programs, as demonstrated in [39]. The binary tree book-keeping methods presented there should be readily extendable to incorporate the particularities of loop corrections. What is missing, of course, is the prescription for the loop corrections themselves. For example, in order to introduce one-loop corrections to the N-leg amplitude, one would, in principle, need the analytic expression for all N-leg one loop integrals with all legs on-shell, as well as all one-loop integrals with $m < N$ legs and $min(m, N-m)$ legs off-shell! These are certainly unknown for $N > 6$. Moreover, there is no generic way to obtain results for loop integrals of increased complexity, for the moment[10]. One could imagine that if such a generic way actually exists, it would be realized by a rearrangement of terms among Feynman diagrams[11], and, in such a case, it might be that the CM algorithm would have to be abandoned. However, this being highly speculative, it is hard to find a compelling reason not to employ the CM algorithm in multi-leg calculations at present times.

[10] See [62] for a lucid review of the available methods.

[11] Like, for example, the one presented in [57], which, though, fails to incorporate loop corrections, at present.

Chapter 3

Monte Carlo and Quasi-Monte Carlo

A method for calculating the error in the Quasi-Monte Carlo method for numerical integration is presented. An estimator for the error is suggested along with a number of modifications that can reduce the computational cost. The relative merit of Quasi-Monte Carlo over the ordinary Monte Carlo integration method is exhibited for a collection of integrands in various dimensions.
The plots that appear in this chapter can also be viewed, in full color detail, at the web address `http://www.ru.nl/imapp/theory/mc_qmc`. A c++ implementation of the error estimator suggested here is also available at the above address, along with the code necessary to generate all Quasi-Monte Carlo pointsets mentioned in this thesis.

3.1 Monte Carlo and Quasi-Monte Carlo

The general N-particle phase space integral, involved in the calculation of a cross-section or decay rate with N outgoing particles, is of the form (see eq. 1.10, 1.11):

$$\int \prod_{k=1}^{N} \left[\frac{d^4 p_k}{(2\pi)^4} \delta(p_k^2 - m^2) \theta(p_k^0) \right] (2\pi)^4 \delta^4(p_A + p_B - \sum_i p_i) \tag{3.1}$$

Explicit analytical expressions for such integrations are restricted to the case of two (or in some cases three) particles in the final state, whereas the modern need for results involves processes with up to ten produced particles. Since the number of (unconstrained) integrations in a process with N particles in the final state is $3N-4$, we see that for a modest $N = 5$ case we end up with an eleven-dimensional integration. Performing that analytically is quite hopeless.

Numerical integration becomes necessary at this point. Geometrically speaking, the integration region is a complicated intersection of four-dimensional hyperboloids. It would be particularly cumbersome and unnatural to try and impose an ordinary lattice structure in such a region, upon which to perform the numerical integration using standard deterministic techniques. But there is a far more important reason to use stochastic methods for such integrals: the error of the stochastic methods far outperforms that of any known deterministic algorithm.

3.1.1 About Monte Carlo and Quasi-Monte Carlo

The name 'Monte Carlo' was given by N.Metropolis [49] to a method of numerical integration based on statistical sampling of the integration volume [1]. Since the late 50's there has been a widespread use of the technique within, but also outside, physics. The importance of Monte Carlo has increased spectacularly following the computational power boom of the beginning of the 80's. Since then a number of computer programs have been developed in the high energy particle physics field, to calculate cross-sections and decay rates, or, alternatively, to generate 'events' for any possible physical process (see [65], [18], [**?**] for big scale libraries that incorporate methods to calculate effects beyond the hard scattering, but also [40],[45],[8] for programs dedicated to matrix element generation).

[1] The method was originally developed by N.Metropolis and S.Ulam under the stimulating influence of J. von Neumann, in the Los Alamos research laboratory, in 1947, for the study of neutron diffusion. E.Fermi had been using similar methods without publishing any results since the early thirties. See [48] for a historical review by Metropolis, where he states that his suggestion on the name Monte Carlo 'was not unrelated to the fact that Stan [Ulam] had an uncle who would borrow money from relatives because he 'just had to go to Monte Carlo' '.

In any method of numerical integration, the integral is estimated by its discrete counterpart, a sum of the value of the integrand over a sequence of points in the integration volume.

$$\int dx\ f(x) \to \frac{1}{N}\sum_{i=1}^{N} f(x_i) \tag{3.2}$$

The defining property of Monte Carlo is the fact that these points are chosen randomly out of a collection of points distributed in an iid[2] uniform way over the integration region[3]. The actual x_i, and, consequently, the value of the function $f(x_i)$ are then considered to be stochastic variables, and the average of $f(x_i)$ - which is seen to coincide with the sum in eq.3.2 - is purported to approach the value of the integral in the limit of large number of points. The higher moments of $f(x_i)$ are related to the integration error as we will soon see. Quasi-Monte Carlo is a variant of Monte Carlo in which the integration points are not chosen independently, but rather with an explicit interdependence so that their overall distribution is 'smoother' or more 'uniform', according to some concrete measure of uniformity to be described below. As a result, the integration error is supposed to be much smaller than in the ordinary Monte Carlo case. The main problem with Quasi-Monte Carlo is the lack of a reliable way to estimate *h*ow much smaller that integration error is.

3.1.2 Visual representation of the 'smoothness' of Quasi-Monte Carlo sets

Despite the fact that uniformity (or 'smoothness') can and will be defined in a fully quantitative way in later sections, it is useful to remark here that what makes Quasi-Monte Carlo sets smoother is their property of non-clustering. To be more precise, the Quasi-Monte Carlo point-sequences are *c*onstructed in such a way as to explicitly avoid clustering. Clustering of points is, on the other hand, an anticipated effect for truly random point-sets, hence our pseudo-random point-sets should also exhibit it[4].

The following figure, fig.3.1.2, shows a scatterplot of a 2-dimensional RANLUX [44] point-sets and one of a 2-dimensional Van de Corput [34] point-sets. RANLUX is the most frequently used pseudo-random sequence, whereas Van de Corput is not the most sophisticated, but definitely one of the easiest to implement Quasi-Monte Carlo sequences. We will make extensive use of both in this thesis.

The clustering of the `RANLUX` set is obvious, when compared with the `Van De Corput` set. This phenomenon is in the heart of the relatively slow convergence of the Monte Carlo estimate to the actual value of the estimated integral. It is by attempting to remedy this clustering problem that Quasi-Monte Carlo hopes to provide a faster converging estimator.

3.1.3 The importance of the errors

In stochastic numerical integration, the main problem is not to obtain a numerical answer[5] for the integral, but rather to ensure that the inherent numerical error is as small as possible, and, on the other hand, to estimate this error as precisely as possible. For integrands with well-known smoothness properties, *a-priori* estimates of the numerical error are possible, but for most practical applications the smoothness properties of the integrand can only be investigated in the course of the integration itself, that is, by repeated numerical evaluation of the integrand.

In stochastic integration methods of the Monte Carlo or Quasi-Monte Carlo types, the integration error is itself an estimate, which contains its own error. That this is not an academic point becomes clear when we realize that the error estimate is routinely used to provide

[2]Here and in the rest of the thesis iid stands for 'independent, identically distributed'.

[3]This ignores the possible interpretation of stratified and importance sampling methods of variance reduction. These can, at any rate, always be formulated in terms of methods using iid uniform integration points.

[4]Or else they are not good pseudo-random sets.

[5]Which is known to be 42, see [1].

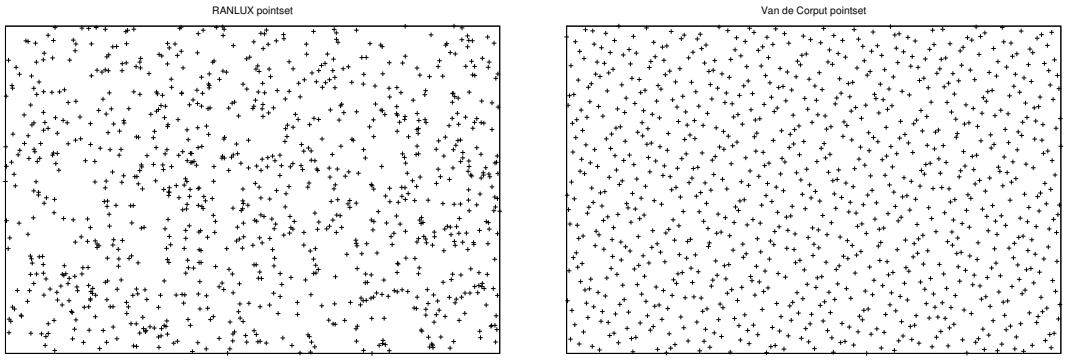


Figure 3.1: 2-dimensional scatterplot of `RANLUX` (left), and `Van Der Corput` (right) point-sets, each containing 1000 points.

confidence levels for the integral estimate (be it based either on Chebyshev or Central-Limit-Theorem, Gaussian rules[6]); and a mis-estimate of the integration error can lead to a serious under- or overestimate of the confidence level. As an example, suppose that the Central Limit Theorem is applicable, so that the integration result is drawn from a Gaussian distribution centered around the true integral value. One standard deviation, as estimated by Monte Carlo, corresponds to a two-sided confidence level of 68%. If the error estimate is off by 50% (admittedly a large value), the actual confidence level may then be anything between 38% and 87%.

From this consideration, we are led to a hierarchy of error estimates: the *first-order* error is that on the integral estimate, while the *second-order* error is the error on the error estimate. This in turn has, of course, its own *third-order* error, and so on. Higher orders than the second one, however, appear to be too academic for practical relevance, but we should like to argue that, in any serious integration problem, the second-order error ought to be included. In what follows we shall discuss the first- and second-order error estimates.

Due to the absence of a Quasi-Monte Carlo error estimator, users of Quasi-Monte Carlo have been estimating the integration error with the classical Monte Carlo formula, as if the point-set was iid. This systematically overestimates the error in any case where the quasi point-set is of any worth. Moreover, no confidence levels can be assigned since the classical estimator does not average to the error made by the quasi, non-iid point-sequence. The purpose of this chapter is to investigate possible estimators for Quasi-Monte Carlo integration taking under consideration the non-iid nature of the underlying point-set[7].

3.2 Monte Carlo estimators

In this section we briefly review the probabilistic theory underlying Monte Carlo integration. For a complete introduction to Monte Carlo methods see the review articles [38] or [68] (where

[6] A confidence level of $X\%$ means that the possibility for a measurement to be further than one standard deviation away from the cited average value is $(100-X)\%$. In other words, there is $X\%$ chance that a measurement would yield a value within one σ from the average. See [38] for details on confidence levels in the Chebyshev or the Gaussian sense.

[7] The opposite direction - re-introducing randomness by reshuffling the points of the Quasi-Monte Carlo sequence in a way that preserves their uniformity properties, thus allowing for the use of a 'classical'-type estimator - has been studied extensively in the literature (see [51] and references therein). Such point-sequences behave better than Monte Carlo sequences and, for integrands with certain properties, as good as Quasi-Monte Carlo sequences. Estimating the error, however, requires the use of a number r of different reshufflings of a point-set with n points, thereby trading off accuracy for knowledge of the error.

an introduction to Quasi-Monte Carlo and quasi-random numbers can be found).

We shall consider integration problems over the d-dimensional unit hypercube $C = [0,1)^d$. The integrand is a function $f(\vec{x})$, which we shall assume real and non-negative, and, of course, integrable over C. We shall define

$$J_m = \int_C f(\vec{x})^m \, d^d\vec{x} \quad , \quad m = 1, 2, 3, \ldots \quad , \tag{3.3}$$

so that J_1 is the required integral. Note that J_m is not necessarily finite for $m \geq 2$. In Monte Carlo we assume N integration points, to be chosen iid from the uniform probability distribution over C. This means that the *point-set* $X = \{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N\}$ on which the integration is based is *assumed* to be a typical member of an ensemble of such point-sets[8], in such a way that the combined probability distribution of the N points over this ensemble is the uniform iid one:

$$P_N(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N) = 1 \quad . \tag{3.4}$$

We shall take the averages over this ensemble. Other assumptions on the underlying ensemble from which the point-set X is believed to be chosen are possible, leading to a different form of P_N. In this, the situation is not different from that encountered in statistical mechanics. The above assumption, however, is the one that is always made in regular Monte Carlo and is justified to some extent by the fact that good-quality (pseudo)random number generators are actually available, allowing us to build ensembles of point-sets X that indeed have the above property (3.4).

Let us assume that a point-set X has been generated, and the values of the integrand $f(\vec{x})$ at all these points have been computed. These we shall denote by $f_j \equiv f(\vec{x}_j)$, $j = 1, 2, \ldots, N$. From these we can compute the discrete analogues of the integrals J_m, which are computable in linear time (that is, time proportional to N):

$$S_m = \sum_{j=1}^{N} (f_j)^m \quad . \tag{3.5}$$

The Monte Carlo estimate of the integral is then

$$E_1 = \frac{1}{N} S_1 \quad . \tag{3.6}$$

The expected value of E_1 over the above ensemble of point-sets is given by

$$\langle E_1 \rangle = \frac{1}{N} \sum_i \langle f_i \rangle = \int_C f(\vec{x}) \, d^d\vec{x} = J_1 \quad , \tag{3.7}$$

which is indeed the required integral: this is the basis for the Monte Carlo method. Its usefulness appears if we compute the variance of E_1:

$$\sigma\,(E_1)^2 = \left\langle E_1^2 \right\rangle - \langle E_1 \rangle^2 = \frac{1}{N} \left(J_2 - J_1^2 \right) \quad . \tag{3.8}$$

[8]Note that the point-set needs only behave *as* if it was random, i.e., the sole demand is that the points are distributed as uniformly as truly random points would be, and that the point-set passes a series of randomness tests in a satisfactory way. Truly random point-sets are never actually used in Monte Carlo, since it is very cumbersome to produce them and impossible to reproduce them. Instead we use point-sets called pseudo-random, designed to imitate truly random ones as good as possible. Of course these point-sets are produced in a fully deterministic way. The absolute source of information on algorithms used to produce pseudo-random sequences remains the classic [42]. A short review of algorithms for generation of pseudo- and quasi- point-sets can be found in [68].

Since this decreases as N^{-1}, the Monte Carlo method actually converges for large N. Note that the leading, $\mathcal{O}(N^0)$, terms of $\langle E_1^2 \rangle$ and $\langle E_1 \rangle^2$ cancel each other: this is a regular phenomenon in variance estimates of this kind[9]. The variance $\sigma\left(E_1\right)^2$ is estimated by the first-order error estimator (also called 'classical' or 'pseudo' estimator in what follows)

$$E_2 = \frac{1}{N^2} S_2 - \frac{1}{N^3} S_1^2 \quad , \tag{3.9}$$

for which we have

$$\langle E_2 \rangle = \sigma\left(E_1\right)^2 + \mathcal{O}(N^{-2}) \quad . \tag{3.10}$$

Since N is usually quite large, at least 10,000 or so, we feel justified in working only to leading order in N. The squared error of E_2 is computed to be, to leading order in N,

$$\sigma\left(E_2\right)^2 = \frac{1}{N^3} \left(J_4 - 4 J_3 J_1 - J_2^2 + 8 J_2 J_1^2 - 4 J_1^4 \right) \quad , \tag{3.11}$$

for which the estimator is

$$E_4 \quad = \quad \frac{1}{N^7} \left(N^3 S_4 - 4 N^2 S_3 S_1 - N^2 S_2^2 + 8 N S_2 S_1^2 - 4 S_1^4 \right) \quad . \tag{3.12}$$

which can also be computed in linear time; we have

$$\langle E_4 \rangle = \sigma\left(E_2\right)^2 + \mathcal{O}(N^{-4}) \quad . \tag{3.13}$$

Some details on the computation of leading-order expectation values of this type, as well as (for purposes of illustration) the form of the third- and fourth-order error estimators, E_8 and E_{16}, respectively, are given in Appendix B.

A final remark is in order here. The Central Limit theorem, which ensures that the error estimate can be used to derive *Gaussian* confidence levels, can also be inferred from the computation of the higher cumulants of the error distribution: we find for the skewness

$$\left\langle \left(E_1 - \langle E_1 \rangle \right)^3 \right\rangle = \frac{1}{N^2} \left(J_3 - 3 J_2 J_1 + 2 J_1^3 \right) \quad , \tag{3.14}$$

and the unnormalized kurtosis:

$$\left\langle \left(E_1 - \langle E_1 \rangle \right)^4 \right\rangle - 3\sigma\left(E_1\right)^2 = \frac{1}{N^3} \left(J_4 - 4 J_3 J_1 - 3 J_2^2 + 12 J_2 J_1^2 - 6 J_1^4 \right) \quad , \tag{3.15}$$

which indicate that the higher cumulants decrease faster than the variance with increasing N; we shall examine this later on for the case of Quasi-Monte Carlo.

3.3 Quasi-Monte Carlo estimators

Multi-point distribution and correlation functions

In contrast to the case of regular Monte Carlo, the technique of Quasi-Monte Carlo relies on point-sets in which the points are *not* chosen iid from the uniform distribution, but rather interdependently. To make this more specific, let us consider a point-set X of N points. For each such a point-set, we may define a *measure of non-uniformity*, called a *discrepancy* or, as in this thesis, a *diaphony*. Its precise definition is presented below: for now, suffice it to demand that there exist a function $D(X)$ of the point-set, which increases with its non-uniformity:

[9] It should be pointed out that what we estimate is the average of the squared error, rather than the error itself, and squaring and averaging do *not* commute. In fact, this is another reason why the second-order estimate is relevant.

$D(X) = 0$ if the point-set is perfectly uniform in all possible respects, an ideal situation that can never be obtained for any finite point-set. The Quasi-Monte Carlo method consists of using point-sets X for which $D(X)$ has some value s which is (very much) smaller than $\langle s \rangle$, the value that may be expected for truly iid uniform ones.

Given that such 'quasi-random' point-sets can be obtained, how does one use them in numerical integration? The obvious issue here is to determine of what ensemble the quasi-random point-set X can be considered to be a 'typical' member. Here, we should like to advocate the viewpoint that, since the main additional property of the quasi-random point-set that distinguishes it from truly random point-sets is its 'anomalously small' discrepancy D, the ensemble ought to consist of those point-sets that are iid uniformly, with the additional constraint that the discrepancy D has the particular value $D(X) = s$ for the actually used point-set[10]. On this premise, the Quasi-Monte Carlo analogue of eq.(3.4) would then be the assumption

$$P(s; \vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N) = \frac{1}{H(s)} \delta(D(X) - s) \ , \tag{3.16}$$

where s is, again, the observed value of the discrepancy of X, on which P_N must now of course depend; and $H(s)$ is the probability density to happen upon a point-sets X with this discrepancy in the regular-Monte Carlo ensemble:

$$H(s) = \int\limits_C \delta(D(X) - s) \, d^d\vec{x}_1 \, d^d\vec{x}_2 \, d^d\vec{x}_N \tag{3.17}$$

The actual computation of $H(s)$ for given definition of the discrepancy is deferred to the next section. What interests us here is the fact that P_N is now no longer simply unity, since that would imply independence of the points in the point-set. Let us therefore write the *multi-point distribution* as

$$P(s; \vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N) = 1 - \frac{1}{N} F(s; \vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N) \ , \tag{3.18}$$

where we have anticipated a factor $1/N$ in the *multi-point correlation* F.

3.3.1 Properties of the correlation function

Since the value of the discrepancy of a given point-set X, should be independent of the order in which the points are generated, $F(s; \vec{x}_1 \ldots \vec{x}_k)$ must be totally symmetric; moreover, we must have

$$F(s; \vec{x}_1, \vec{x}_2, \ldots, \vec{x}_K) = \int\limits_C F(s; \vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k, \vec{x}_{k+1}) \, d^d\vec{x}_{k+1} \ , \tag{3.19}$$

which is not as trivial as it might seem since the value of the discrepancy, s, is based on the full N points and not on the smaller set of k or $k+1$ points. Finally, for the Quasi-Monte Carlo integral to be unbiased, we must have

$$P(s; \vec{x}_1) = 1 \ , \tag{3.20}$$

so that

$$\int\limits_C F(s; \vec{x}_1, \vec{x}_2) \, d^d\vec{x}_2 = 0 \ . \tag{3.21}$$

These remain, of course, to be proven and we shall do so in the next section, for a particular choice of discrepancy. Moreover, we shall show there that the multi-point correlation F_N is, to leading order in $1/N$, made up from two-point correlations F_2:

$$F(s; \vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k) = \sum_{1 \le m < n \le k} F(s; \vec{x}_m, \vec{x}_n) \ . \tag{3.22}$$

[10] We do not examine the possible alternative that the point-sets in the ensemble must have discrepancy *in the neighborhood* of the observed value s; this amounts to the distinction between the micro-canonical and the canonical ensemble in statistical mechanics.

This establishes the properties of our ensemble of point-sets X on which, in our view, the Quasi-Monte Carlo estimates ought to be based.

3.3.2 Estimators

We shall indicate the 'Quasi-Monte Carlo' nature of the estimators by the superscript (q). The first estimator is that of the integral:

$$E_1^{(q)} = \frac{1}{N} \sum f_j \ . \tag{3.23}$$

Here, and in the rest of this section, the sums will run from 1 to N. Denoting by the subscript (q) averages with respect to the 'quasi-random' ensemble discussed above, we then have

$$\left\langle E_1^{(q)} \right\rangle_{(q)} = \int\limits_C \ f(\vec{x}) \ P(s;\vec{x}) \ d^d\vec{x} = J_1 \ , \tag{3.24}$$

as before: owing to the fact that the one-point distribution is uniform, the Quasi-Monte Carlo estimate is indeed as unbiased as the Monte Carlo one. The distinction between the two methods appears in the first-order error estimate. We have

$$\sigma \left(E_1^{(q)} \right)^2_{(q)} = \frac{1}{N} \left(J_2 - J_1^2 - \int f_1 f_2 F_{12} \right) \quad + \quad \mathcal{O}\left(\frac{1}{N^2} \right) \ . \tag{3.25}$$

where we have adopted the straightforward convention for integrals

$$\int f_1 f_2 F_{12} = \int\limits_C \ f(\vec{x}_1) \ f(\vec{x}_2) \ F(s;\vec{x}_1,\vec{x}_2) \ d^d\vec{x}_1 \ d^d\vec{x}_2 \ , \tag{3.26}$$

etc. As before, we shall insouciantly neglect terms that are sub-leading in $1/N$. The advantage of the Quasi-Monte Carlo method is now clear: if we can ensure that $\alpha_{12} > 1$ 'where it counts', that is, generally, when $\vec{x}_1$ and $\vec{x}_2$ are 'close' in some sense, then the Quasi-Monte Carlo error will be smaller than the Monte Carlo one. A good Quasi-Monte Carlo point-set, therefore, is one in which the points 'repel' each other to some extent.

The first-order error estimate is now simply

$$E_2^{(q)} = \frac{1}{N^2} \sum f_i^2 - \frac{1}{N^3} \sum f_i f_j - \frac{1}{N^3} \sum f_i f_j F_{ij} \ . \tag{3.27}$$

It is simple to show that, indeed

$$\left\langle E_2^{(q)} \right\rangle_{(q)} = \sigma \left(E_1^{(q)} \right)^2_{(q)} + \mathcal{O}(N^{-2}) \ ; \tag{3.28}$$

however, *evaluating* $E_2^{(q)}$ is less trivial since it is not obvious how to do this in time linear in N. We shall discuss this later.

Defining

$$\alpha_{ij} = 1 + F_{ij} \tag{3.29}$$

the variance of the estimator $E_2^{(q)}$ can be evaluated to

$$\begin{aligned} \sigma \left(E_2^{(q)} \right)^2 \ = \ & \frac{1}{N^3} \left(\int f_i^4 - 4 \int f_i^3 f_j \alpha_{ij} - \int f_i^2 f_j^2 \alpha_{ij} \right. \\ & +4 \int f_i^2 f_k f_l \alpha_{ik} \alpha_{kl} + 4 \int f_i^2 f_k f_l \alpha_{ik} \alpha_{il} \\ & \left. -4 \int f_i f_j f_k f_l \alpha_{ij} \alpha_{jk} \alpha_{kl} \right) + \mathcal{O}(N^{-4}) \ , \end{aligned} \tag{3.30}$$

for which the corresponding estimator (to leading order) is

$$\begin{aligned} E_4^{(q)} &= \frac{1}{N^7}\Big(N^3\sum f_i^4 - 4N^2\sum f_i^3 f_j\alpha_{ij} - N^2\sum f_i^2 f_j^2\alpha_{ij} \\ &\quad +4N\sum f_i^2 f_k f_l\alpha_{ik}\alpha_{kl} + 4N\sum f_i^2 f_k f_l\alpha_{ik}\alpha_{il} \\ &\quad -4\sum f_i f_j f_k f_l\alpha_{ij}\alpha_{jk}\alpha_{kl}\Big) \quad . \end{aligned} \tag{3.31}$$

The details of this computation are discussed in the Appendix.

It goes without saying that the substitution $F_{ij} \to 0$ (or $\alpha_{ij} \to 1$) will reduce all the Quasi-Monte Carlo results to the regular Monte Carlo ones.

We can now see why the 'classical' estimator eq.(3.9) overestimates the error. Under the quasi distribution P_2 of eq.(3.18) the classical estimator averages to

$$\begin{aligned} \langle E_2\rangle_{(q)} &= \left\langle \frac{1}{N^2}\sum_i f_i^2 - \frac{1}{N^3}\sum_{i,j} f_i f_j \right\rangle_{(q)} \\ &= \frac{1}{N}(J_2 - J_1^2) - \frac{1}{N^2}\int f_x f_y F_{x,y} + \mathcal{O}(\frac{1}{N^2}) \end{aligned} \tag{3.32}$$

The term involving the correlator is suppressed by $\frac{1}{N}$, which shows that E_2 averages to something different than the variance of E_1 under the quasi distribution. Moreover, we will show in section 3.4.5 that[11] the integral of the suppressed term is strictly positive for any point-set that is better than a truly random one. So E_2 omits a strictly negative term when estimating the error.

While it is true that the estimator eq.(3.27) averages to a quantity whose leading order in N is equal to the leading order of $\sigma\left(E_2^{(q)}\right)^2$, it suffers from the following disagreeable property: for a constant integrand, while the first two terms vanish identically, the third approaches zero asymptotically from negative values. This leads to a negative squared error for all practical purposes. Although this is not disastrous *per se* it indicates the reason for the appearance of negative errors also for non-constant integrands, as will become apparent once we have a concrete expression for the correlation function. It is, thus, desirable to obtain an estimator that vanishes identically for constant functions. This is achieved by

$$E_2^{(q_2)} = \frac{1}{N^{\underline{2}}}\sum_i f_i^2 - \frac{1}{NN^{\underline{2}}}\hat{\sum_{i,j}} f_i f_j - \frac{1}{NN^{\underline{4}}}\hat{\sum_{i,j,k,l}} f_i f_j (F_{i,j} - F_{i,k} - F_{l,j} + F_{l,k}) \tag{3.33}$$

where the $\hat{\Sigma}_{i,j\ldots}$ denotes a sum with all indices different, and $F_{i,j} \equiv F_2(s;\vec{x}_i,\vec{x}_j)$. This quantity averages to

$$\left\langle E_2^{(q_2)}\right\rangle = \frac{1}{N}(J_2 - J_1^2) - \frac{1}{N}\int dxdydzdw\; f(x)f(y)\left[F_{x,y} - F_{x,w} - F_{z,y} + F_{z,w}\right] \tag{3.34}$$

which equals the leading part of $\sigma\left(E_2^{(q)}\right)^2$ thanks to eq.(3.21). It is easy to check that the estimator of eq.(3.33) vanishes identically for a constant integrand and any N, thanks to the antisymmetry property of the quadruple sum.

3.3.3 Cumulants of E_1

As a final remark, we may also investigate the cumulants of the Quasi-Monte Carlo estimator E_1. We write the expansion of the correlation function F_k over $1/N$ as

$$F_k(s;\vec{x}_1,\ldots,\vec{x}_k) \equiv F_k^{(1)} + \frac{1}{N}F_k^{(2)} + \frac{1}{N^2}F_k^{(3)} + \ldots \tag{3.35}$$

[11] Under fairly general conditions for the function $f(x)$.

and define

$$\mathcal{M}^{(a)}_{i_1,\ldots,i_k} \equiv \int f(\vec{x}_1)^{i_1} \ldots f(\vec{x}_k)^{i_k} F^{(a)}_k(s;\vec{x}_1,\ldots,\vec{x}_k) \tag{3.36}$$

It is evident that if eq.(3.22) holds, we have

$$\mathcal{M}^{(1)}_{1,1,\ldots,1} = \frac{k^{\underline{2}}}{2} J_1^{k-2} \mathcal{M}^{(1)}_{1,1} \tag{3.37}$$

The cumulants are defined as

$$c_n = \left\langle \left(E^{(q)}_1 - \left\langle E^{(q)}_1 \right\rangle_{(q)} \right)^n \right\rangle_{(q)} \tag{3.38}$$

The variance of E_1 is then

$$c_2 = \frac{1}{N}(J_2 - J_1^2 - \mathcal{M}^{(1)}_{1,1}) + O(\frac{1}{N^2}) \tag{3.39}$$

The skewness is

$$c_3 = \frac{1}{N^2}(J_3 - 3J_1J_2 + 2J_1^3 - 3\mathcal{M}^{(1)}_{1,2} + 3J_1\mathcal{M}^{(2)}_{1,1} + 6J_1\mathcal{M}^{(1)}_{1,1} - \mathcal{M}^{(2)}_{1,1,1}) + O(\frac{1}{N^3}) \tag{3.40}$$

The unnormalized kurtosis is

$$c_4 - 3c_2^2 = \frac{1}{N^2}(-\mathcal{M}^{(2)}_{1,1,1,1} - 3(\mathcal{M}^{(1)}_{1,1})^2 + 4J_1\mathcal{M}^{(2)}_{1,1,1} - 6J_1^2\mathcal{M}^{(2)}_{1,1}) + O(\frac{1}{N^3}) \tag{3.41}$$

The above results indicate that a correlation function that satisfies the property of eq.(3.22) leads to a distribution whose skewness decreases faster with N than does the variance, but when it comes to the kurtosis (and higher cumulants), additional properties regarding the next-to-leading order expression for F (denoted above by $\mathcal{M}^{(2)}_{i_1,\ldots,i_k}$) are needed to secure Gaussian cumulants[12]. These properties hold whenever the saddle point approximation of eq.(3.73-3.74) is valid. In such cases one expects Gaussian confidence levels for the Quasi-Monte Carlo estimator E_1.

[12]Approach to a Gaussian distribution,for iid random variables, would require $c_n/(c_2)^{n/2}$ to approach 0 for large N.

3.4 Multi-point distributions with diaphonies

3.4.1 Diaphony

We consider a point set X with N elements, given in C. The non-uniformity of the point set X can be described by its *diaphony*[13]:

$$D(X) = \frac{1}{N} \sum_{j,k=1}^{N} \beta(\vec{x}_j, \vec{x}_k) \ , \tag{3.42}$$

with

$$\begin{aligned} \beta(\vec{x}_j, \vec{x}_k) &= \hat{\sum_{\vec{n}}} \sigma_{\vec{n}}^2 \, e_{\vec{n}}(\vec{x}_j) \bar{e}_{\vec{n}}(\vec{x}_k) \ , \\ e_{\vec{n}}(\vec{x}) &= \exp(2i\pi \ \vec{n} \cdot \vec{x}) \ . \end{aligned} \tag{3.43}$$

Here, the vectors $\vec{n} = (n_1, n_2, \ldots, n_d)$ form the integer lattice, and the hat denotes the sum over all $\vec{n}$ except $\vec{n} = \vec{0}$. We may also write

$$D(X) = \frac{1}{N} \hat{\sum_{\vec{n}}} \sigma_{\vec{n}}^2 \left| \sum_{j=1}^{N} e_{\vec{n}}(\vec{x}_j) \right|^2 \ , \tag{3.44}$$

so that we recognize the diaphony as a measure of how well the various Fourier modes are integrated by the point set X. The diaphony is therefore seen to be related to the 'spectral test', well-known in the field of random-number generator testing. For the *mode strengths* $\sigma_{\vec{n}}^2$ we have

$$\sigma_{\vec{n}}^2 \geq 0 \quad , \quad \hat{\sum_{\vec{n}}} \sigma_{\vec{n}}^2 = 1 \ . \tag{3.45}$$

The latter convention simply establishes the overall normalization of D. The advantage of this diaphony over, say, the usual (star)discrepancy is the fact that it is translation-invariant:

$$\beta(\vec{x}_j, \vec{x}_k) = \beta(\vec{x}_j - \vec{x}_k) \ , \tag{3.46}$$

so that point sets X and X' that differ only by a translation (modulo 1) have the same non-uniformity: the diaphony is actually defined on the hyper-torus rather than on the hypercube. Also, the diaphony is *tadpole-free*[14]:

$$\int_C \beta(\vec{x}, \vec{x}') \ d^d\vec{x} = 0 \ . \tag{3.47}$$

and its beta-function has the attractive property

$$\int_C \beta(\vec{x}, \vec{x}) \ d^d\vec{x} = 1 \ . \tag{3.48}$$

Moreover, we shall use $\sigma_{\vec{n}}^2$ such that $\sigma_{\vec{n}}^2 = \sigma_{\vec{n}'}^2$ if the two lattice vectors $\vec{n}$ and $\vec{n}'$ differ only by a permutation of their components. Thus, X and X' will also have the same non-uniformity if they differ by a global permutation of the coordinates of the points.

[13] some of the concepts of this section have also been discussed in [35] and [36].

[14] The term tadpole-free refers to the diagrammatic approach of calculating the diaphony of a random point-set. Its meaning will be clear in section 3.4.4.

3.4.2 The expected diaphony for an iid point-sets

In the case of an iid point-sets, the diaphony becomes a stochastic variable whose expected value is

$$\begin{aligned}\langle D(\{x_i\})\rangle &= \frac{1}{N}\hat{\sum_{\vec{n}}}\sigma_{\vec{n}}^2\sum_{i,j=1}^{N}\langle e^{2i\pi\vec{n}(\vec{x}_i-\vec{x}_j)}\rangle = \\ &= \frac{1}{N}\hat{\sum_{\vec{n}}}\sigma_{\vec{n}}^2\sum_{i=j}\int dx e^0 + \frac{1}{N}\hat{\sum_{\vec{n}}}\sigma_{\vec{n}}^2\sum_{i\neq j}\int dx\, dy\, e^{2i\pi\vec{n}\vec{x}}e^{-2i\pi\vec{n}\vec{y}} = \\ &= \frac{1}{N}\sum_{i=j}\hat{\sum_{\vec{n}}}\sigma_{\vec{n}}^2\int dx e^0 = 1 \end{aligned} \tag{3.49}$$

We see that the diaphony is normalized in such a way that it averages to one for a truely random iid point-set. Any quasi-random point-set will have to do better than that, and should, ideally, have a diaphony whose value approaches zero asymptotically.

3.4.3 Some numerical results

In this section the behavior of a specific diaphony is presented[15], for three point sequences, as the number of points N increases.

The diaphony is defined by eq.(3.44) with

$$\sigma_{\vec{n}} = Ke^{-\lambda\vec{n}^2} \quad K^{-1} = \sum e^{-\lambda\vec{n}^2} \quad \lambda = 0.1 \tag{3.50}$$

The reason for experimenting with this definition lies in the factorizing property of the $\sigma_{\vec{n}}$. Due to K^{-1} being related to Jacobi theta functions, we call this the 'Jacobi diaphony'. We will be using this diaphony in most of what follows.

In this thesis we will be using three sequences that we will be calling `RANLUX`, `Van Der Corput` and `Niederreiter`. `RANLUX` is a pseudo-random point sequence generated by the `RANLUX` algorithm (see [44]) with luxury level equal to 3. `Van Der Corput` is a quasi-random sequence generated by an implementation of the algorithm by Halton that generalizes to many dimensions an older algorithm by Van der Corput (see [34]) with prime bases $2, 3, 5, 7, 11, \ldots$. Finally `Niederreiter` is another, optimal[16] quasi-random sequence based on the algorithm in [50]. In particular, we follow the choices of [13] and construct the sequence in whichever base is optimal for the current dimension (see [50]). The construction of the Niederreiter sequence is an interesting topic on its own, so we have devoted appendix A to details on the relevant algorithm. .

The definition of the diaphony, eq.3.44, includes a sum over modes which, in principle, extends to infinity. In practice, of course, we will truncate the sum after a number of terms. In the plots presented below we have set an a priori maximum in the number of modes, called M. The computer finds,then, the maximum squared length among the first M modes [17] and removes all modes whose squared length exceeds that maximum. This way the number of modes included is kept within strict limits, and, if a mode with square length x is included, so are all the other modes whose square length is x.

In the plots of fig.3.2, fig.3.3, the diaphony of the `Niederreiter` sequence in particular, but also that of the `van der Corput` sequence, exhibits a large variation in relatively small intervals of N. As the number of points N approaches certain critical values the diaphony reaches very small levels, only to return to its 'cruising' values a few points later. To avoid cluttering the

[15] The plots that appear in this chapter can also be viewed, in full color detail, at the web address *http* : *//www.ru.nl/imapp/theory/mc_qmc*.

[16] Optimal in a sense described in [50] and [13].

[17] The first M modes are selected out of the square lattice.

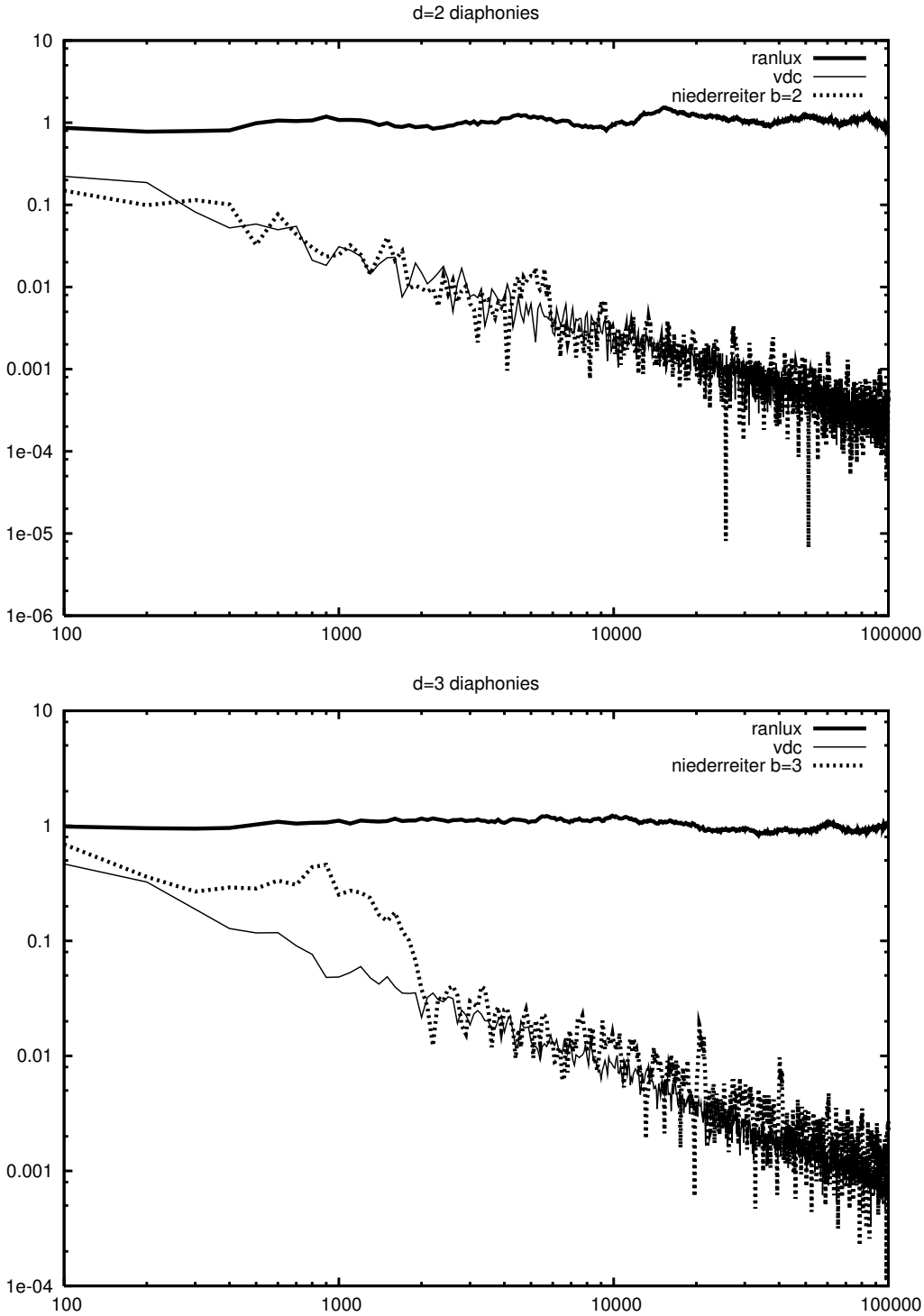


Figure 3.2: The diaphony of `RANLUX` (thick line), `Van Der Corput` (thin line) and `Niederreiter` (dotted line), for D=2,3.

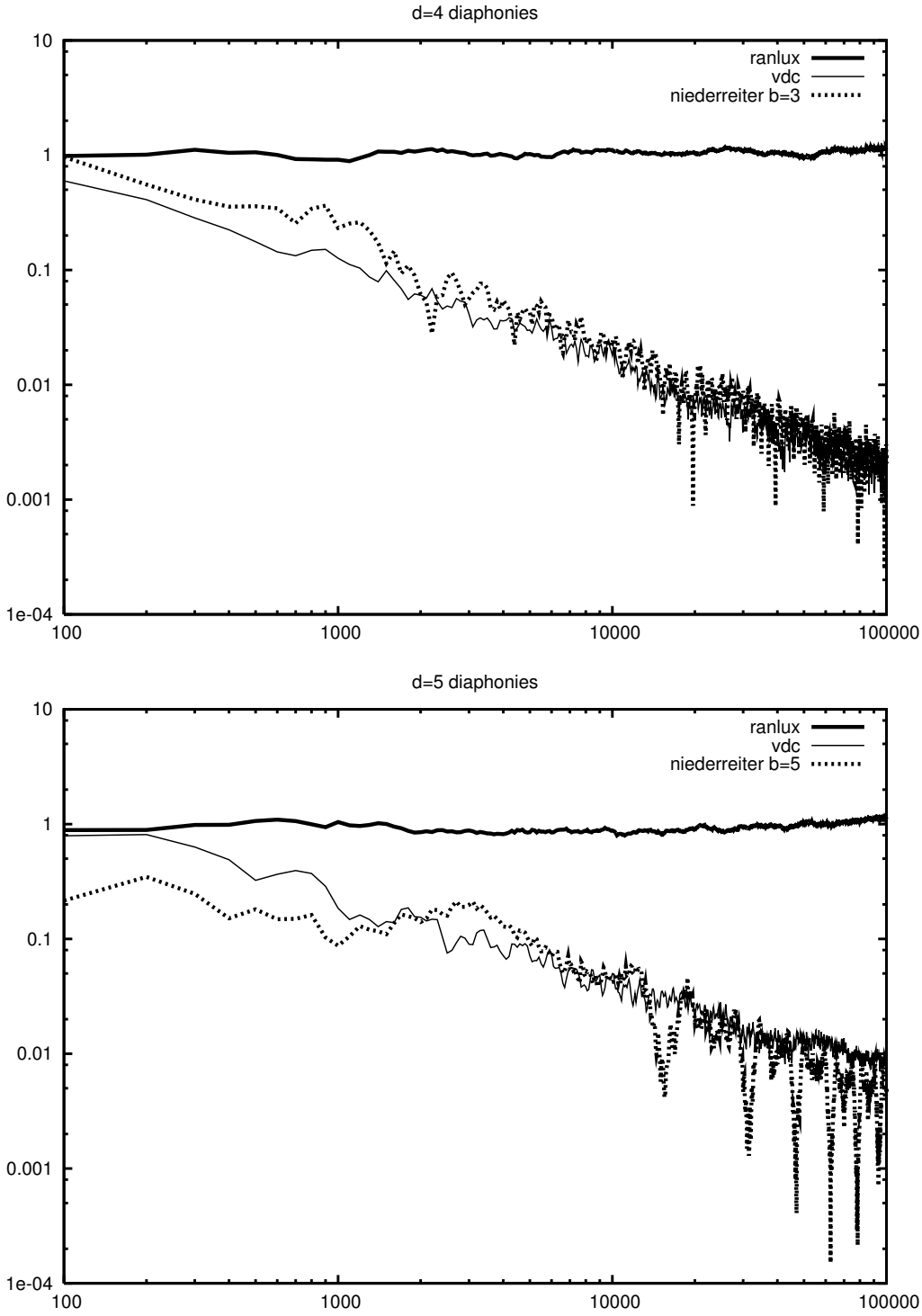


Figure 3.3: The diaphony of `RANLUX` (thick line), `Van Der Corput` (thin line) and `Niederreiter` (dotted line), for D=4,5.

plots we present here the diaphony averaged in packs of 500 points without information on the minimum or maximum value found in each pack. The minimum values for each pack, that correspond to exceptional point configurations, are very interesting on their own but do not affect the present study.

The diaphony of the `RANLUX` sequence is seen to oscillate around 1, as expected. Moreover the behavior of the `Niederreiter` sequence improves with the number of dimensions when compared with crude `Van der Corput`, an encouraging hint for higher dimensions.

3.4.4 Generating function

We shall now compute a $1/N$ approximation to the function

$$G_p(z) = \Big\langle \exp(zD(X)) \Big\rangle_{\vec{x}_{p+1},\vec{x}_{p+2},\ldots,\vec{x}_N} \quad , \tag{3.51}$$

where we have indicated that the points $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_p$ are kept fixed while the remaining $N-p$ points are integrated over. $G_p(z)$ therefore still depends on $\vec{x}_1, \ldots, \vec{x}_p$.

This function generates the moments of the diaphony with respect to the $N-p$-point probability distribution when p of the N points are already fixed.

Expanding the exponential we have

$$\begin{aligned} G_p(z) &= \sum_{r=0}^{\infty} \frac{z^r}{r!} \langle D^r \rangle_{\vec{x}_{p+1},\vec{x}_{p+2},\ldots,\vec{x}_N} \\ &= \sum_{r=0}^{\infty} \frac{z^r}{r!} \frac{1}{N^r} \sum_{i_1,j_1,i_2,j_2,\ldots,i_r,j_r} \langle \beta(i_1,j_1)\ldots\beta(i_r,j_r) \rangle_{\vec{x}_{p+1},\vec{x}_{p+2},\ldots,\vec{x}_N} \\ &= \sum_{r=0}^{\infty} \frac{z^r}{r!} \frac{1}{N^r} \sum_{i_1,j_1,i_2,j_2,\ldots,i_r,j_r} \int d\vec{x}_{p+1}\, d\vec{x}_{p+2} \ldots\, d\vec{x}_N \beta(i_1,j_1)\ldots\beta(i_r,j_r) \end{aligned} \tag{3.52}$$

This multiple sum is most easily achieved using a diagrammatic approach, which has been introduced in [35]. We shall indicate with crosses those points that are kept fixed[18] (with an implied sum over them, from 1 to p), and with dots ('beads') those points that are integrated over (again, with an implied sum running from $p+1$ to N). The function β is indicated by a solid line. As the simplest examples, then, we have

$$\text{if } p = N\text{:} \quad \frac{1}{N} \times\!\!\longrightarrow\!\!\times = \frac{1}{N} \sum_{j,k=1}^{N} \beta(\vec{x}_j - \vec{x}_k) = D(X) \quad , \tag{3.53}$$

and

$$\text{if } p = 0\text{:} \quad \langle D(X) \rangle_{\vec{x}_1,\ldots,\vec{x}_N} = \beta(0) = \bigcirc = 1 \quad . \tag{3.54}$$

Other examples are

$$\begin{aligned} \bigcirc &= \int_C \beta(\vec{x}_1 - \vec{x}_2)^2\, d^d\vec{x}_1\, d^d\vec{x}_2 \quad , \\ \text{(two beads, two lines)} &= \int_C \beta(\vec{x}_1 - \vec{x}_2)\beta(\vec{x}_2 - \vec{x}_3)\beta(\vec{x}_3 - \vec{x}_1)\, d^d\vec{x}_1\, d^d\vec{x}_2\, d^d\vec{x}_3 \quad , \end{aligned} \tag{3.55}$$

and so on: a general closed loop with precisely n beads will be denoted by (n). Note that, since, the functions $e_{\vec{n}}(\vec{x})$ form an orthonormal (and even a complete) set, we have

$$\text{(n)} = \hat{\sum_{\vec{n}}} \left(\sigma^2_{\vec{n}}\right)^n \quad . \tag{3.56}$$

[18] Also called 'external' in the jargon.

We can now simply write out all possible (connected and disconnected) diagrams where every solid line ends in a cross or a bead, and apply the following Feynman rules:

1. A factor $2z/N$ for every β line (where the factor 2 arises from the two possible orientations);
2. A factor $(N-p)^{\underline{q}}$ for every diagram (or product of diagrams) that contains precisely q beads[19];
3. In addition, the usual symmetry factors arising from equivalent lines and vertices, and from the repetition of identical (sub)diagrams.

We shall compute $G_p(z)$ including terms of order 1 and those of order $1/N$. Note that

$$(N-p)^{\underline{q}} = N^q \left(1 - \frac{pq}{N} - \frac{q(q-1)}{2N}\right) + \mathcal{O}(N^{-2}) \tag{3.57}$$

as long as $N \gg pq, q^2$. In the following we shall always assume this.

First, we consider contributions without any crosses or nontrivial vertices. A general term in this class is given by

$$\frac{(N-p)^{\underline{Q}}}{N^Q} \frac{1}{r_1!} \Big(z \bigcirc\Big)^{r_1} \frac{1}{r_2!} \left(z^2 \bigcirc\right)^{r_2} \frac{1}{r_3!} \left(\frac{4z^3}{3} \bigcirc\right)^{r_3} \cdots \ ,$$

where

$$Q = r_1 + 2r_2 + 3r_3 + \cdots \ ; \tag{3.58}$$

up to order $1/N^2$, this contribution to the generating function can therefore be written as

$$\begin{aligned} G_p^{(1)}(z) &= \left(1 - \frac{pz}{N}\frac{\partial}{\partial z} - \frac{z^2}{2N}\frac{\partial^2}{\partial z^2}\right) \sum_{\{r\}} \prod_{n\geq 1} \frac{1}{r_n!} \left(\frac{(2z)^n}{2n} \boxed{\mathrm{n}}\right)^{r_n} \\ &= \left(1 - \frac{pz}{N}\frac{\partial}{\partial z} - \frac{z^2}{2N}\frac{\partial^2}{\partial z^2}\right) G^{(0)}(z) \ , \\ G^{(0)}(z) &= \exp\left(-\frac{1}{2}\hat{\sum_{\vec{n}}} \log\left(1 - 2z\sigma_{\vec{n}}^2\right)\right) \ . \end{aligned} \tag{3.59}$$

Up to $1/N^2$, one diagram with a four-point vertex may be present: a generic contribution of this type is

$$\frac{(N-p)^{\underline{Q+m+1+m_2+1}}}{N^{Q+m_1+m_2+2}} \left(\frac{(2z)^{m_1+m_2+2}}{8} \ \mathrm{m}_1 \bigcirc\!\!\bigcirc \mathrm{m}_2\right)$$
$$\times \frac{1}{r_1!} \Big(z \bigcirc\Big)^{r_1} \frac{1}{r_2!} \left(z^2 \bigcirc\right)^{r_2} \frac{1}{r_3!} \left(\frac{4z^3}{3} \bigcirc\right)^{r_3} \cdots \ ,$$

where $m_{1,2}$ denote the number of beads on each loop, excluding the one on the four-vertex. Let us define

$$\phi(z; \vec{x}_j - \vec{x}_k) = \hat{\sum_{\vec{n}}} \frac{2z\sigma_{\vec{n}}^2}{1 - 2z\sigma_{\vec{n}}^2} e_{\vec{n}}(\vec{x}_j) \bar{e}_{\vec{n}}(\vec{x}_k) \ ; \tag{3.60}$$

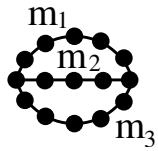
then, this contribution can be written as

$$G_p^{(2)}(z) = \frac{1}{8N} \phi(z;0)^2 G^{(0)}(z) \ . \tag{3.61}$$

[19] The 'falling power' is defined as $a^{\underline{b}} = a!/(a-b)! = a(a-1)(a-2)\cdots(a-b+1)$.

Note that the lemniscate graph is actually equal to the product of two closed loops: this is a consequence of the translational invariance of the diaphony. A generic contribution containing two three-vertices is

$$\frac{(N-p)^{\underline{Q+m_1+m_2+m_3+2}}}{N^{Q+m_1+m_2+m_3+3}} \left(\frac{(2z)^{m_1+m_2+m_3+3}}{12} \right.$$



$$\left. \times \frac{1}{r_1!} \left(z \right)^{r_1} \frac{1}{r_2!} \left(z^2 \right)^{r_2} \frac{1}{r_3!} \left(\frac{4z^3}{3} \right)^{r_3} \cdots \right) ,$$

so that this contribution to the generating function reads

$$G_p^{(3)}(z) = \frac{1}{12N} G^{(0)}(z) \int\limits_C \phi(z;\vec{x})^3 \, d^d\vec{x} \ . \tag{3.62}$$

The diagrams with crosses have the generic contribution

$$\frac{(N-p)^{\underline{Q+m}}}{N^{Q+m+1}} \left(z(2z)^m \right.$$

$$\left. \times \frac{1}{r_1!} \left(z \right)^{r_1} \frac{1}{r_2!} \left(z^2 \right)^{r_2} \frac{1}{r_3!} \left(\frac{4z^3}{3} \right)^{r_3} \cdots \right) ,$$

leading to

$$\begin{aligned} G_p^{(4)}(z) &= \frac{1}{2N} G^{(0)}(z) \sum_{j,k=1}^{p} \phi(z;\vec{x}_j - \vec{x}_k) \\ &= \frac{1}{N} G^{(0)}(z) \left(\frac{p}{2} \phi(z;0)^2 + \sum_{1 \le j < k \le p} \phi(z;\vec{x}_j - \vec{x}_k) \right) \ , \end{aligned} \tag{3.63}$$

where we have singled out the contributions with $j = k$. All other possible diagrams either vanish because of translational invariance and tadpole-freedom, or are of order $1/N^2$ or lower.

The final result for the generating function up to order $1/N^2$ is therefore

$$\begin{aligned} G_p(z) &= G^{(0)}(z) \left(1 - \frac{1}{4N} \int\limits_C \phi(z;\vec{x})^2 \, d^d\vec{x} + \frac{1}{12N} \int\limits_C \phi(z;\vec{x})^3 \, d^d\vec{x} \right. \\ &\qquad \left. + \frac{1}{N} \sum_{1 \le j < k \le p} \phi(z;\vec{x}_j - \vec{x}_k) \right) \ . \end{aligned} \tag{3.64}$$

with

$$G^{(0)}(z) = \exp\left(-\frac{1}{2} \hat{\sum_{\vec{n}}} \log\left(1 - 2z\sigma_{\vec{n}}^2 \right) \right) \tag{3.65}$$

and

$$\phi(z;\vec{x}_j - \vec{x}_k) = \hat{\sum_{\vec{n}}} \frac{2z\sigma_{\vec{n}}^2}{1 - 2z\sigma_{\vec{n}}^2} e_{\vec{n}}(\vec{x}_j) \bar{e}_{\vec{n}}(\vec{x}_k) \ ; \tag{3.66}$$

Note that the term in $G^{(1)}(z)$ containing p cancels precisely against that in $G^{(4)}(z)$, so that the only reference to p is in the last term in brackets in eq.(3.64), and indeed we have

$$\int\limits_C G_p(z) \, d^d\vec{x}_p = G_{p-1}(z) \ . \tag{3.67}$$

In Appendix B.4 we give the result for the higher order ($O(\frac{1}{N^2})$) term in G_p. There are 25 terms that contribute but only three of them include p. The condition 3.67 still holds.

3.4.5 Multi-point distribution by Laplace transform

From the generating function, we can recover the actual probability distributions. As discussed above, let $H(s)$ be the probability that the point set X has diaphony equal to s, that is, $D(X) = s$. The underlying ensemble of point sets is that of sets of N iid uniformly distributed points, *i.e.* the same ensemble underlying the usual Monte Carlo error estimates. Then, we have

$$\begin{aligned} H(s) &= \int_C d^d\vec{x}_1\, d^d\vec{x}_2 \cdots d^d\vec{x}_N\, \delta(D(X) - s) \\ &= \frac{1}{2i\pi} \int_{-i\infty}^{+i\infty} e^{-zs}\, G_0(z)\, dz \ , \end{aligned} \tag{3.68}$$

where the integration contour runs to the left of all the singularities of $G_0(z)$. The latter is given by eq.3.64 for $p = 0$.

The multi-point distribution for p points averaged over all point sets X with diaphony s, is given by

$$\begin{aligned} P(s;\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_p) &= \frac{1}{H(s)} R(s;\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_p) \ , \\ R(s;\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_p) &= \frac{1}{2i\pi} \int_{-i\infty}^{+i\infty} e^{-zs}\, G_p(z)\, dz \ . \end{aligned} \tag{3.69}$$

Since we write the deviation from uniformity of the multi-point distribution as

$$P(s;\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_p) = 1 - \frac{1}{N} F(s;\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_p) \ , \tag{3.70}$$

we see that the multi-point correlation $F(s;\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_p)$ is, up to $O(\frac{1}{N})$, as claimed, built up from two-point correlators[20]: for $p \geq 3$,

$$F(s;\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_p) = F(s;\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_{p-1}) + \sum_{j=1}^{p-1} F(s;\vec{x}_j,\vec{x}_p) \ , \tag{3.71}$$

so that the p-point correlator is simply the sum of all $p(p-1)/2$ 2-point correlators.

In the approximation used, the sub-leading terms in $H(s)$ are actually irrelevant, and we may write

$$\begin{aligned} H(s) &\approx \frac{1}{2i\pi} \int_{-i\infty}^{+i\infty} \exp(\psi(s;z))\, dz \ , \\ \psi(s;z) &= -sz - \frac{1}{2} \hat{\sum_{\vec{n}}} \log\left(1 - 2z\sigma^2_{\vec{n}}\right) \ , \\ F(s;\vec{x}_1,\vec{x}_2) &= \frac{-1}{2\pi i H(s)} \int_{-i\infty}^{+i\infty} \exp(\psi(s;z))\phi(z;\vec{x}_1 - \vec{x}_2)\, dz \ . \end{aligned} \tag{3.72}$$

Except in the very simplest cases[21], a complete evaluation of eq.(3.72) is nontrivial. A simplification arises if s is much smaller than its expectation value 1 (which is anyway the aim in

[20] This doesn't hold for the next order in $\frac{1}{N}$ as seen in Appendix B. Terms like the one of eq.(B.18), that don't factorize, appear for $p \geq 3$.

[21] See section 3.6.3.

quasi-Monte Carlo), or if the Gaussian limit is applicable, namely when the number of modes with non-negligible $\sigma^2_{\vec{n}}$ becomes large in such a way that no single mode dominates. In practice, this happens when the dimensionality of C becomes large. Fortunately, these are precisely the situations of interest. The position of the saddle point for $H(s)$, $\hat{z}$, is given by

$$\hat{\sum_{\vec{n}}} \frac{\sigma^2_{\vec{n}}}{1 - 2\hat{z}\sigma^2_{\vec{n}}} = s \quad . \tag{3.73}$$

For $s \ll 1$, therefore, $\hat{z}$ is large and negative. Since to first order the same saddle point may be used for R_2, we find the attractive result

$$F(s; \vec{x}_1, \vec{x}_2) \approx \hat{\sum_{\vec{n}}} \omega_{\vec{n}}\, e_{\vec{n}}(\vec{x}_1)\, \bar{e}_{\vec{n}}(\vec{x}_2) \quad , \quad \omega_{\vec{n}} = \frac{-2\hat{z}\sigma^2_{\vec{n}}}{1 - 2\hat{z}\sigma^2_{\vec{n}}} \quad . \tag{3.74}$$

The formulae (3.73) and (3.74) suffice, in our approximation, to compute all the multi-point correlations.

We finish this section with the following observation. Suppose that F_2 is given as a function of $\vec{x}_1, \vec{x}_2$. By Fourier integration we can then compute the $\omega_{\vec{n}}$. The assumption that the saddle-point approximation is valid, together with the normalization condition $\sum \sigma^2_{\vec{n}} = 1$, then allows us to write

$$\hat{z} = -\frac{1}{2}\hat{\sum_{\vec{n}}} \frac{\omega_{\vec{n}}}{1 - \omega_{\vec{n}}} \quad , \quad \sigma^2_{\vec{n}} = -\frac{1}{2\hat{z}} \frac{\omega_{\vec{n}}}{1 - \omega_{\vec{n}}} \quad , \quad s = \hat{\sum_{\vec{n}}} \sigma^2_{\vec{n}}(1 - \omega_{\vec{n}}) \quad . \tag{3.75}$$

We see that F_2 not only determines the *form* of the diaphony, but in addition also its *value*.

3.5 Application of Quasi-Monte Carlo estimators

3.5.1 The mechanism behind error reduction

After the above preliminaries we can now examine the mechanism by which Quasi-Monte Carlo can outdo Monte Carlo. We shall assume the saddle-point approximation to be valid. For $s < 1$, we then have $\hat{z} < 0$. All the $\omega_{\vec{n}}$ are positive (see eq.3.74), and as $\hat{z} \to -\infty$ they approach unity from below (although for $|\vec{n}| \to \infty$ they must always, of course, go to zero). Now notice that the set of functions $e_{\vec{n}}(\vec{x})$ is complete, that is,

$$\sum_{\vec{n}} e_{\vec{n}}(\vec{x}_1)\, \bar{e}_{\vec{n}}(\vec{x}_2) = \delta^d(\vec{x}_1 - \vec{x}_2) \quad . \tag{3.76}$$

This allows us to write the variance of the Monte Carlo error as

$$\sigma\,(E_1)^2 = \frac{1}{N} \hat{\sum_{\vec{n}}} \left| \int_C f(\vec{x})\, e_{\vec{n}}(\vec{x})\, d^d\vec{x} \right|^2 \quad , \tag{3.77}$$

where the contribution from the zero mode $\vec{n} = 0$ is canceled by the J_1^2 term. For Quasi-Monte Carlo on the other hand, we find

$$\sigma\left(E_1^{(q)}\right)^2_{(q)} = \frac{1}{N} \hat{\sum_{\vec{n}}} (1 - \omega_{\vec{n}}) \left| \int_C f(\vec{x})\, e_{\vec{n}}(\vec{x})\, d^d\vec{x} \right|^2 \quad . \tag{3.78}$$

We see that those modes $\vec{n}$ for which $\omega_{\vec{n}}$ is positive tend to lead to an error reduction. In the saddle-point approximation, therefore, *any* value $0 < s < 1$ will lead to a decreased error with respect to standard Monte Carlo. On the other hand, since

$$0 < \hat{z} < \min_{\vec{n}} \frac{1}{2\sigma_{\vec{n}}^2} \quad \text{for} \quad s > 1 \quad , \tag{3.79}$$

large values of the diaphony will actually lead to an *increase* in the error. Note that in the above we have only used the fact that the $e_{\vec{n}}$ form a *complete, orthonormal* set of functions: therefore, the error-reduction result holds for a much wider class of discrepancies than just the diaphonies discussed in this thesis.

3.5.2 Estimators analyzed

We can now arrive at an estimator for the Quasi-Monte Carlo error. The simplest form is obtained by inserting eq.(3.74) in the equation for E_2, eq.(3.27):

$$E_2^{(q)} = \frac{1}{N^2} \sum f_i^2 - \frac{1}{N^3} (\sum f_i)^2 - \frac{1}{N^3} \sum_{\hat{\vec{n}}} \omega_{\vec{n}} |\sum_i f_i e_{\vec{n}}(x_i)|^2 \tag{3.80}$$

with

$$\omega_{\vec{n}} = \frac{-2\hat{z}\sigma_{\vec{n}}^2}{1 - 2\hat{z}\sigma_{\vec{n}}^2} \tag{3.81}$$

We are still free to choose the exact form of the weights $\sigma_{\vec{n}}^2$ at will, under the constraints of eq.(3.45). Our choice is the so called Jacobi weights[22]

$$\sigma_{\vec{n}}^2 = K e^{-\lambda \vec{n}^2} \tag{3.82}$$

[22] Due to their convenient factorizing property.

with

$$K^{-1} = \hat{\sum}_{\vec{n}} e^{-\lambda \vec{n}^2} \tag{3.83}$$

The parameter λ controls the 'sensitivity' of the diaphony: as $\lambda \to 0$ we get $\sigma_{\vec{n}} \to 1$ for every mode which corresponds to a super-sensitive diaphony, useless for practical purposes, while as $\lambda \to \infty$ only the modes with $\vec{n}^2 = 1$ contribute making the diaphony fairly non-sensitive. We choose $\lambda = 0.1$. Other values of λ, within a 'reasonable range' do not alter, in practice, the numerical value of $E_2^{(q)}$, as shown in section 3.6.1.

It is easy to see that the estimator averages (to leading order in N) in a positive definite quantity [23]. This still leaves open the possibility for a negative error estimate, particularly for relatively smooth functions where the cancelation between the two sums of the pseudo estimate are large leading to a small error. The source of the negative error effect is clear in the case of a constant function. Then

$$f(x) = C \Rightarrow E_2^{(q)} = -\frac{1}{N^3} C^2 \sum_{\vec{n}} \omega_{\vec{n}} \sum_{i,j} u_{\vec{n}}(x_i) \bar{u}_{\vec{n}}(x_j) \tag{3.84}$$

and the point sum of every Fourier mode can be anything from 0 (when the points are spread evenly enough to produce complete cancelations for all the included modes) to N^2 (when all the points are on top of each other). The average of this sum is N (for truly random points), but for Quasi-Monte Carlo points we expect that this sum will be significantly smaller than that. For non-constant functions similar effects can be expected, apart from the fact that the first two terms of $E_2^{(q)}$ do not cancel anymore. Thus, we expect negative squared errors for higher modes or small number of points, and this is what has been observed in a number of plots. Unfortunately there is no way to predict precisely when, as N increases, the estimator gets a useful, positive value. One could resort to the error of $E_2^{(q)}$, but that is cubic in the number of modes (see eq.(3.31)) and hence prohibitively expensive in realistic calculations.

The way out of this is the estimator of eq.(3.33) which can be written in a form with unrestricted sums as follows:

$$\begin{aligned} E_2^{(q)} &= \frac{1}{N^{\underline{2}}} S_2 - \frac{1}{N N^{\underline{2}}} S_1^2 - \frac{(N-1)^{\underline{3}}}{N N^{\underline{4}}} \sum_{\vec{n}} \omega_{\vec{n}} |W_{\vec{n}}|^2 \\ &+ \frac{(N-1)^{\underline{3}}}{N N^{\underline{4}}} S_2 \sum_{\vec{n}} \omega_{\vec{n}} + \frac{N-1}{N N^{\underline{4}}} \sum_{\vec{n}} \omega_n (2 S_1 \Re\left\{W_{\vec{n}} \bar{U}_{\vec{n}}\right\} - 2\Re\left\{U_{\vec{n}} \bar{Q}_{\vec{n}}\right\}) \\ &- \frac{1}{N N^{\underline{4}}} \sum_{\vec{n}} \omega_{\vec{n}} (N - 2 + |U_{\vec{n}}|^2)(S_2 - S_1^2) \end{aligned} \tag{3.85}$$

where

$$\begin{aligned} U_{\vec{n}} &\equiv \sum_i u_{\vec{n}}(x_i) \\ W_{\vec{n}} &\equiv \sum_i u_{\vec{n}}(x_i) f(x_i) \\ Q_{\vec{n}} &\equiv \sum_i u_{\vec{n}}(x_i) f^2(x_i) \\ S_2 &\equiv \sum_i f_i^2 \\ S_1 &\equiv \sum_i f_i \end{aligned} \tag{3.86}$$

[23] It averages to eq.(3.78) which is positive definite as long as $s < 1$.

It is identically zero for a constant function, as can be easily checked, and averages to the leading order of the squared variance of E_1. The correction terms are of higher than leading order in N, but that does *not* mean that we have selectively included some next-to-leading-order corrections to the variance. The correction terms above are such that the next-to-leading-order terms vanish on the average.

In practice the infinite sum over modes in both estimators has to be truncated. This should not be perceived as an approximation of any kind. It amounts to a redefinition of the diaphony. Looking at eq.(3.73) we see that as the value of s becomes small the saddle point becomes quickly large and negative: $\hat{z} \ll 0$. Then $-2\hat{z}\sigma_{\vec{n}}^2 \to \infty$ for low modes and $-2\hat{z}\sigma_{\vec{n}}^2 \to 0$ for higher modes, when $\sigma_{\vec{n}}^2/|\hat{z}| \to 0$. We can, thus, safely neglect these higher modes in the estimator. As long as the value of the diaphony is small, which is in any case the goal in Quasi-Monte Carlo , the profile of $\omega_{\vec{n}}$ depends only on the choice of λ, which, as said, also regulates the sensitivity of the diaphony. We see therefore that the estimator inherits the sensitivity of the diaphony in a direct way.

It is worth noting that the factorized form of the β-function in the diaphony definition is directly responsible for the fact that the two estimators are now of complexity $N \times M$ (with M the number of modes) instead of quadratic in N. This is a desirable achievement as long as $M \leq N$, which we shall always assume to be the case.

3.5.3 Numerical results

In the following we will present a number of plots[24] that show how both the 'classical' and the quasi error estimates[25] behave as a function of the number of points N. In the process we will use the three types of point sequences defined in section 3.4.3.

A number of test functions were used for integrands. They consist of a subset of the test functions used by Schlier in [60], along with a Gaussian function with dimension-dependent width. We have

$$TF13 : f(\vec{x}) = \prod_{k=1}^{D} \frac{|4x_k - 2| + k}{1 + k} \tag{3.87}$$

which averages to $J_1 = 1$. This test function is especially tailored for a Van der Corput sequence, since in $D = 1$ it is perfectly integrated by such a sequence with base 2.

$$TF2 : f(\vec{x}) = \prod_{k=1}^{D} kcos(kx_k) \tag{3.88}$$

which averages to $J_1 = \prod_k sin(k)$. This function should be difficult to integrate in high dimensions.

$$TF4 : f(\vec{x}) = \sum_{k=1}^{D} \prod_{j=1}^{k} x_j \tag{3.89}$$

which averages to $J_1 = 1 - \frac{1}{2^D}$. It is chosen as a simple example of a function that is not a product of single-variable functions.

A Gaussian with fixed width suffers from a rapid decrease, in higher dimensions, of the region of the integration volume where the function is non-zero, making the integration cumbersome

[24]The plots that appear in this chapter can also be viewed, in full color detail, at the web address `http://www.ru.nl/imapp/theory/mc_qmc`.

[25]The 'classical' or 'pseudo' estimator, E_2, is the one of eq.(3.9), constructed on the assumption that the points are iid. By 'quasi' estimator, $E_2^{(q)}$, we mean the 'improved' estimator eq.(3.85).

D	# of modes
1	6
2	44
3	250
4	1256
5	5182

D	# of modes
1	4
2	20
3	56
4	136
5	332

Table 3.1: Number of modes with $\vec{n}^2 \leq 15$ (left) and $\vec{n}^2 \leq 5$ (right)

(the higher the dimension, the more points are needed and inter-dimensional comparison is difficult). To avoid this we use instead

$$TF6 : f(\vec{x}) = \prod_{i=1}^{D} \sum_{n_i=-\infty}^{\infty} \frac{e^{-(x_i - x_{0i} + n_i)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} \tag{3.90}$$

which is a product of superpositions of a Gaussian and its tails outside the $[0, 1]$ interval. We wish to keep the variance of this function independent of the number of dimensions, so we define σ such that

$$\frac{1}{2\sigma} \sum_{m=-\infty}^{\infty} e^{-m^2/4\sigma^2} = (1 + V)^{1/D}\sqrt{\pi} \tag{3.91}$$

where in practice it suffices to keep the first couple of terms in the sum. The function averages to $J_1 = 1$ and spreads as the number of dimension grows ($\sigma \to \infty$ as $D \to \infty$).

In the following plots the error and its estimates as functions of the number of points N are shown in a double logarithmic scale.

The 'classical' error estimate is presented, along with three versions of quasi error estimators, $E_2^{q5}, E_2^{q10}, E_2^{q15}$. The superscript next to q denotes the squared length of the highest modes included in the sums of eq.(3.85). Thus E_2^{q10} includes[26] modes with $\vec{n}^2 \leq 10$. In table 3.1 we give the number of modes with $\vec{n}^2 \leq 15$, and $\vec{n} \leq 5$ for different dimensions. It is evident that the number of modes grows rapidly with the dimensionality.

The real error made is included for comparison. The data were collected at a point-per-point basis up to $N = 10^5$. In the plots we have included the average value of each error for successive subsets of 500 points, suppressing any information on minimum or maximum values in the subset[27].

All integrations are performed in the unit hypercube $[0, 1]^D$. The dimensionality varies from 2 to 6.

[26] Please note that the square length of a mode is the sum of the squares of D integers. So for $D = 2$, for example, the modes present are those with square equal to $1, 2, 4, 5, 9, 10, 13, 16, 17, \ldots$ and, thus, $E_2^{(q15)}$ actually contains modes with squared length up to 13.

[27] The real error (in particular) fluctuates a lot as the quasi sets complete their successive cycles of low diaphony, but knowledge of the specific point where the error minimizes is of course not available a priori.

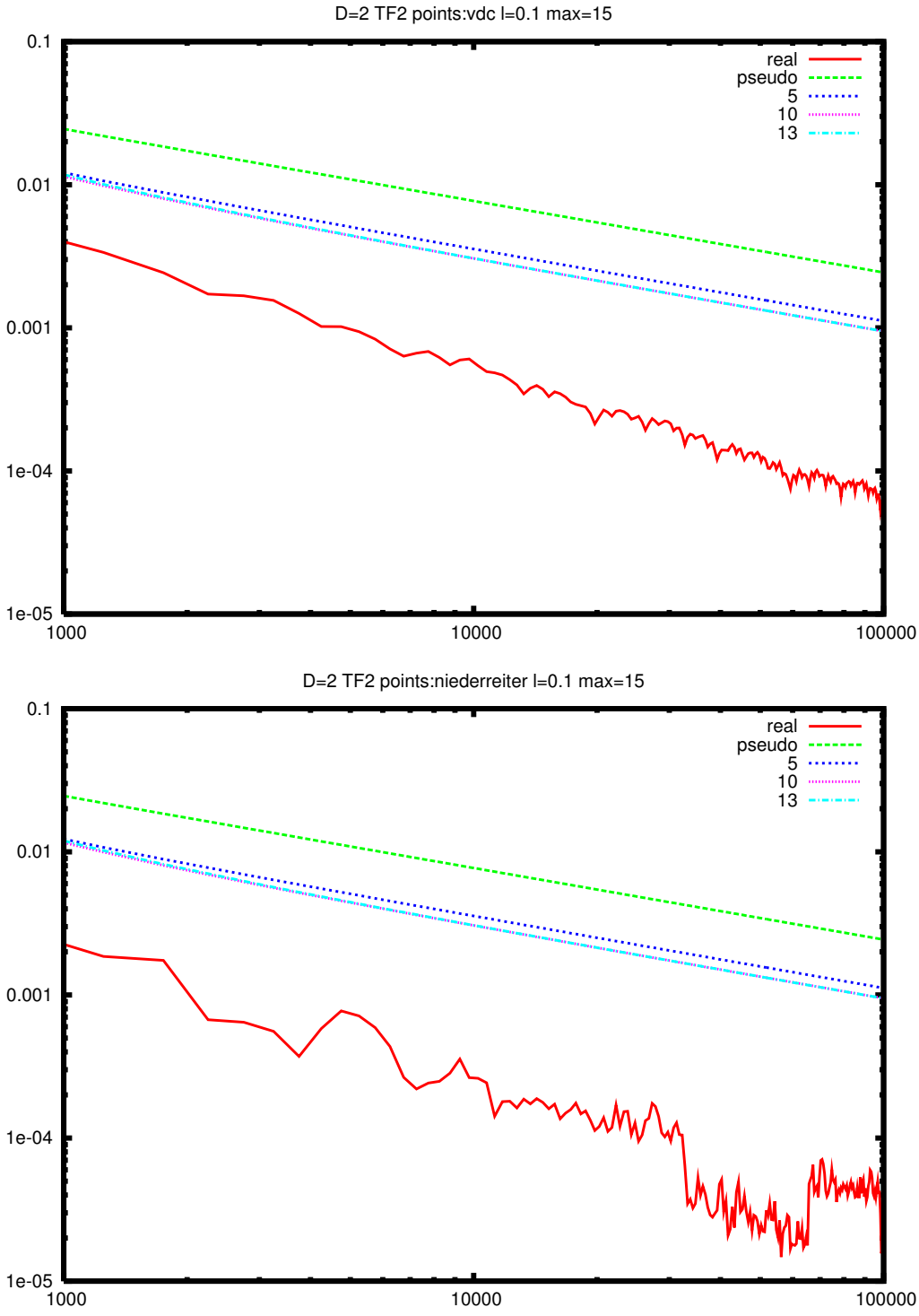


Figure 3.4: TF2, d=2 log-plot using a `Van der Corput` sequence and a `Niederreiter` sequence. The classical error estimator is far off the real error whereas the quasi estimators are approaching the real error as more modes are added to the sum. The need for more modes is, however, obvious, in both plots.

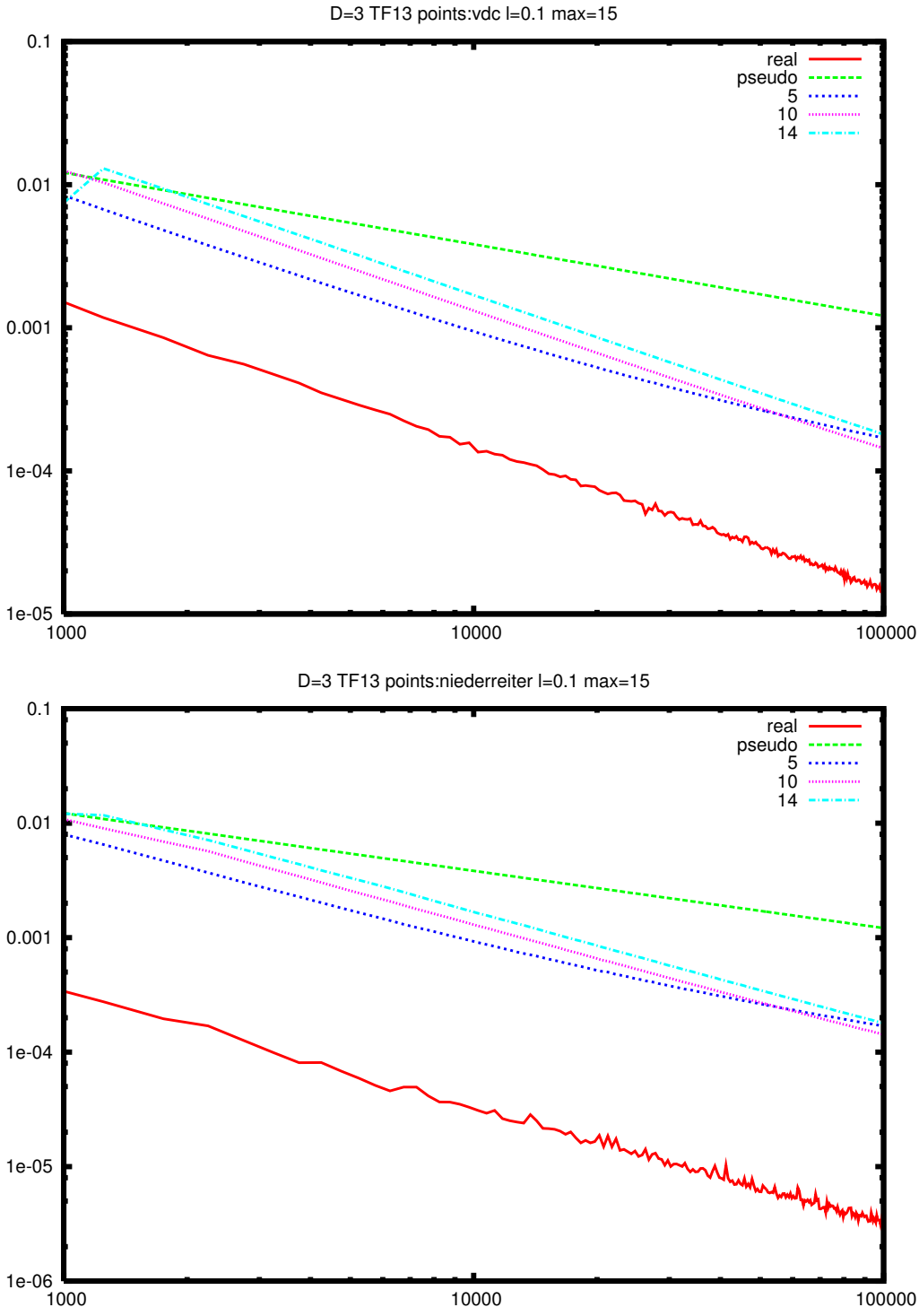


Figure 3.5: TF13, d=3 log-plot using a `Van der Corput` sequence and a `Niederreiter` sequence. The quasi estimators follow the error with the appropriate N-dependence contrary to the pseudo estimator. Note that the E^{q14} is in this case worse than E^{q10} or E^{q5} for all $N \leq 100000$. The higher modes converge slower to their average value, but the cross-over point is not known in advance and it is function-dependent.

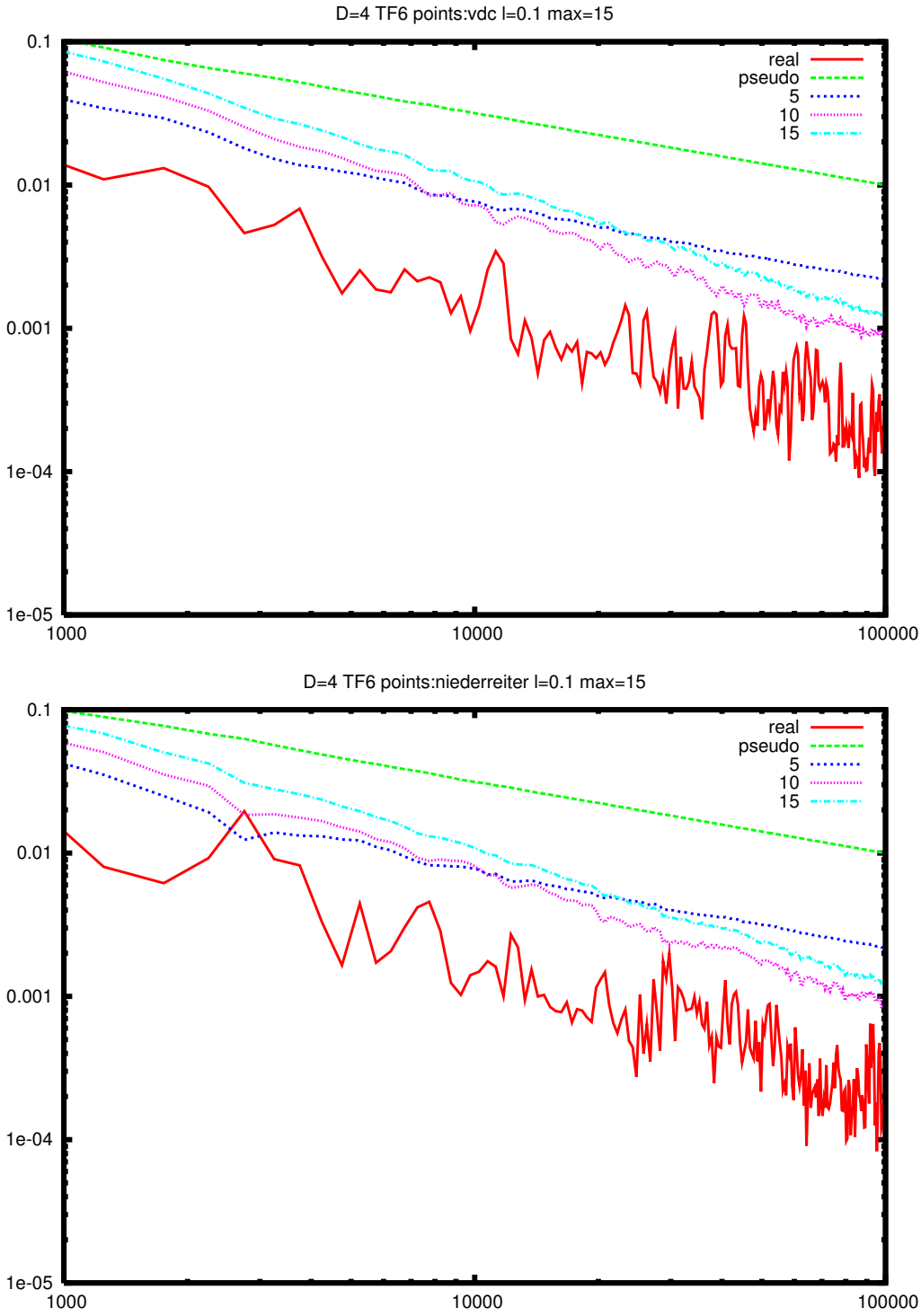


Figure 3.6: TF6, d=4 log-plot using a `Van der Corput` sequence and a `Niederreiter` sequence. The quasi estimators approximate well the error. Moreover we see here a clearer instance of the crossover of higher modes in large N mentioned in the previous figure.

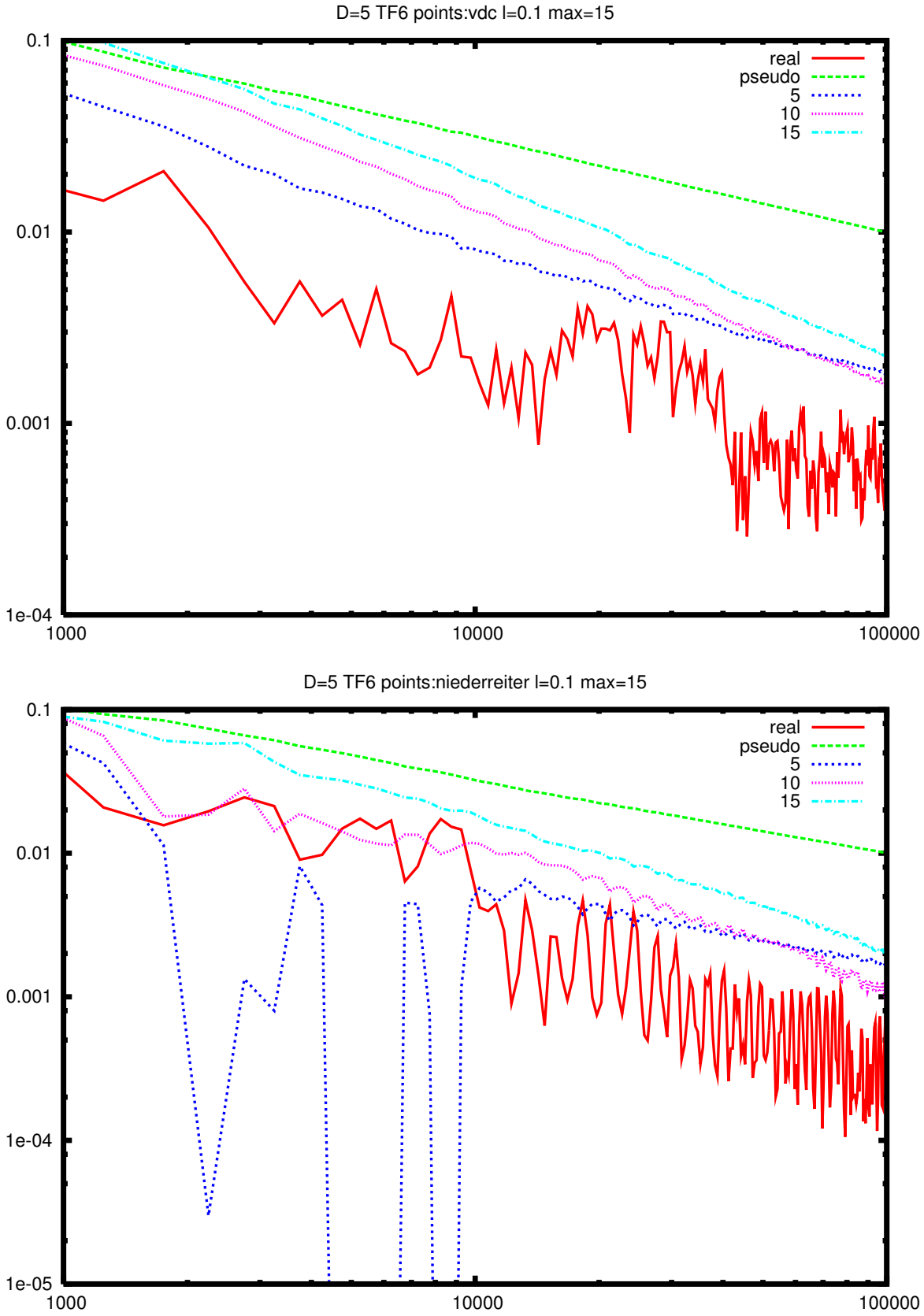


Figure 3.7: TF6, d=5 log-plot using a `Van der Corput` sequence and a `Niederreiter` sequence. The use of the improved estimator, eq.(3.85), reduces the probability of a negative error square estimate but, naturally, it doesn't remove it altogether. The plot on the right demonstrates this effect. As expected, the estimator returns to positive values and stabilizes as the number of points increases and the estimator converges to its average value.

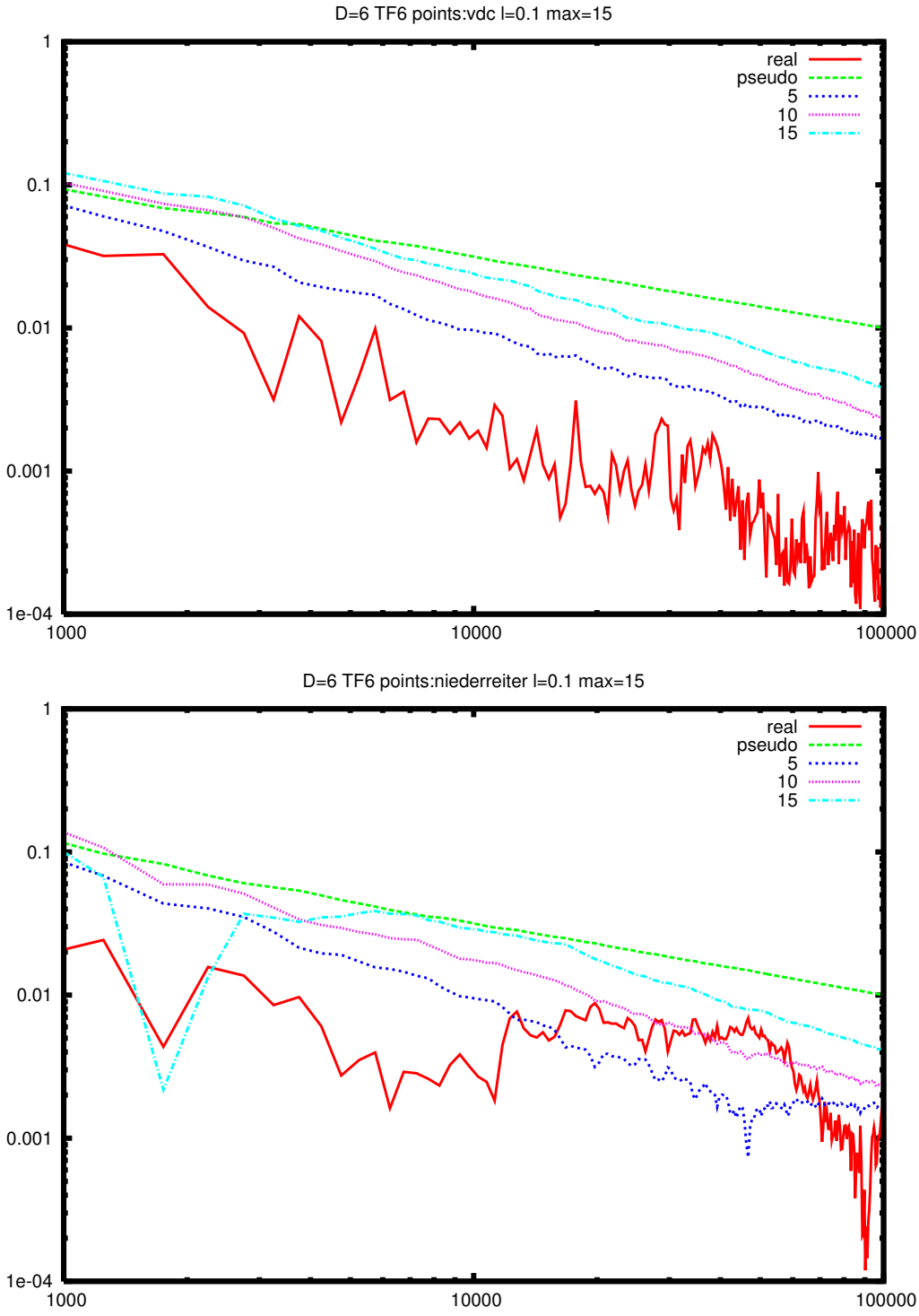


Figure 3.8: TF6, d=6 log-plot using a `Van der Corput` sequence and a `Niederreiter` sequence. In this case the estimators describe very well the real error made in the integration.

3.6 Alternative approaches

3.6.1 Raising the value of λ in the Jacobi diaphony

In general the real Quasi-Monte Carlo error is approached by including more and more modes in the estimator sum. At the same time, by including higher modes, one increases the error on this estimate (the error on E_2) because one attempts to estimate by Monte Carlo means the integral $\int f(\vec{x})e_{\vec{n}}(\vec{x})$ which will fluctuate vigorously for higher modes.

One might then attempt to raise the value of λ, thus decreasing the number of active modes (that give an appreciably non-zero $\omega_{\vec{n}}$)). This would of course reduce the sensitivity of the diaphony, artificially lowering its value. Improvement in the error estimate originating from higher modes would be lost but the contribution of the modes close to the origin (which are the ones included) would be relatively enhanced, as can be seen from the behavior of the weights $\omega_{\vec{n}}$ (see eq.(3.74)) .

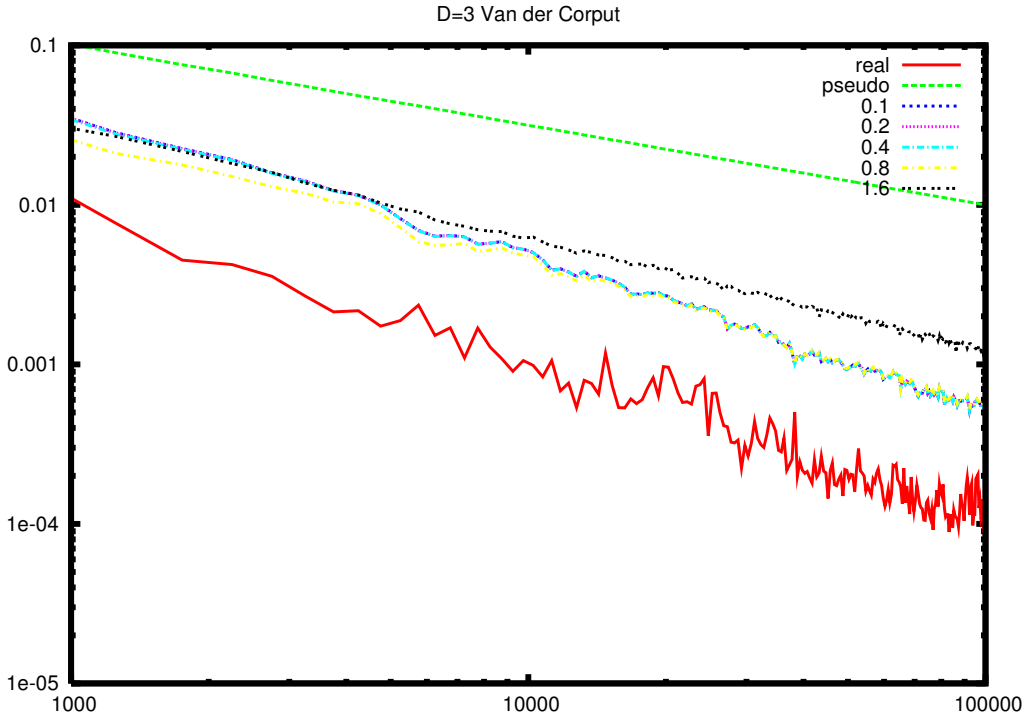


Figure 3.9: TF6, d=4 log-plot using the `Van der Corput` sequence. E^{q15} is shown for different values of λ indicated in the key, along with the real error and the classical estimate. Average values of all quantities for sets of 500 points are shown in each case. The value of λ doesn't alter the estimator, as long as that value stays within a specific range. We see that, in this case, the value $\lambda = 1.6$ is out of the safe range.

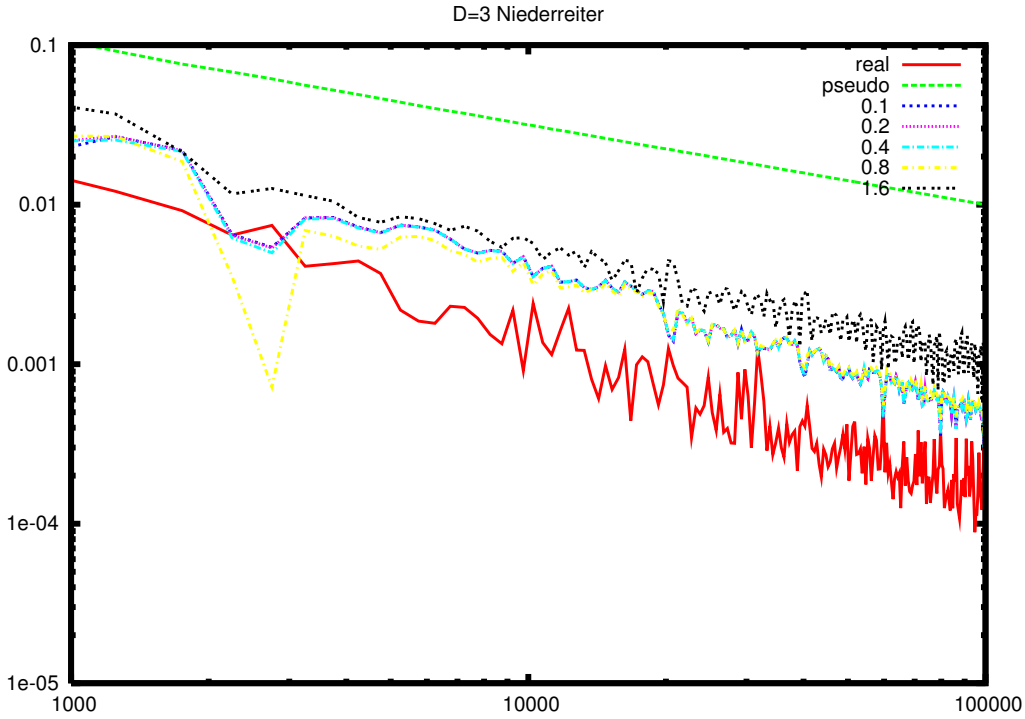


Figure 3.10: TF6, d=4 log-plot using the `Niederreiter` sequence. E^{q15} is shown for different values of λ indicated in the key, along with the real error and the classical estimate. Average values of all quantities for sets of 500 points are shown in each case. The value of λ doesn't alter the estimator, as long as that value stays within a specific range. We see that, in this case, the value $\lambda = 1.6$ is out of the safe range.

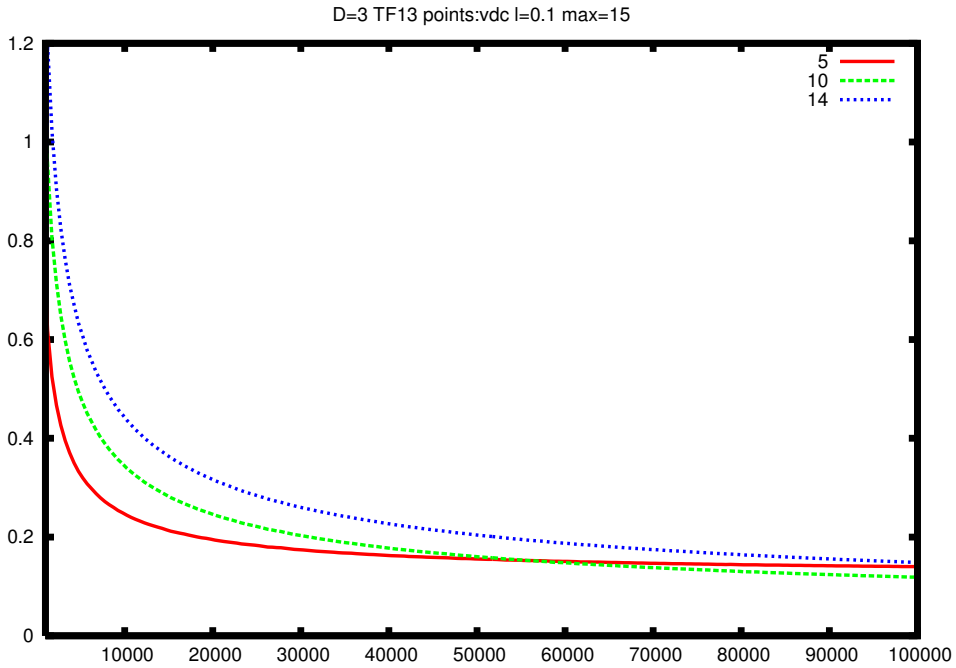


Figure 3.11: The ratio of different quasi estimators with the classical estimator for $d = 3$ and TF13. `Van der Corput` point-sets were used.

3.6.2 Monitored estimator

The estimators E_2^{q5} , E_2^{q10} and E_2^{q15} are not always proportional to the classical estimate, and, in some cases they decreases quite a bit faster with N than the classical estimate does. They never decrease slower than the classical estimate, though, and one can use that as follows. One monitors the ratio of E_2^{q15}, for example, to the 'classical' error estimate, and after a certain point[28], the 'classical' error is only estimated and multiplied with that ratio. This is a purely linear algorithm and therefore very fast. Caution has to be exercised, though, in the way the critical ratio is chosen, in order to avoid configurations where the estimators acquire a very low value for some exceptional value of N.

This approach relies heavily on the, frequently false, assumption that the quasi and classical estimators have the same dependence on N. If this is not so, the new estimate is conservative. One has, thus, the option to trade accuracy for CPU time.

The plots of fig.3.12 show the ratio of E_2^{q5} , E_2^{q10} and E_2^{q15} with E_2 for two particular cases.

3.6.3 The box approximation

There is a choice for the diaphony that allows us to perform the integrals of eq.(3.72) without resorting to the saddle point approximation. That choice is

$$\sigma_{\vec{n}}^2 = \frac{1}{M} \prod_{\mu=1..d} \theta(n^\mu \leq m) \tag{3.92}$$

for some arbitrary m. The normalization, eq.(3.45), determines $M = (2m+1)^D - 1$.

This diaphony includes only a finite number of modes, all of which are equally weighted. It can be seen as an approximation to the Jacobi diaphony since for small λ the latter gives $\sigma_{\vec{n}} \approx 1$ for $|\vec{n}| \leq n_c$ and $\sigma_{\vec{n}} \approx 0$ for $|\vec{n}| \geq n_c$ where n_c is determined implicitly by the value of

[28] which depends on the resources of the user.

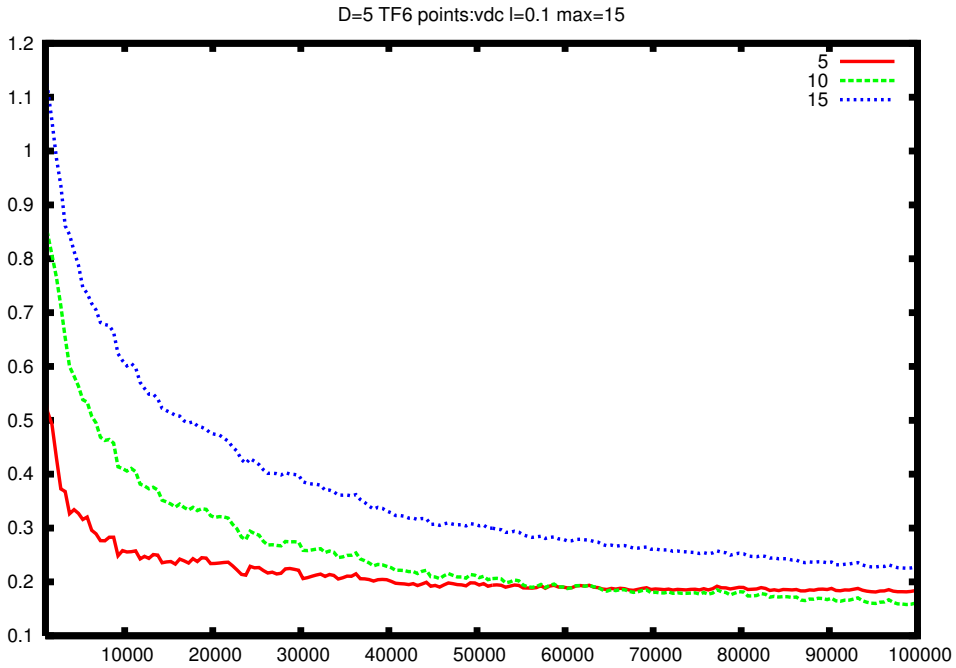


Figure 3.12: The ratio of different quasi estimators with the classical estimator for $d = 5$ TF6. `Van der Corput` point-sets were used.

the Jacobi diaphony. The diaphony can be evaluated as a quadratic function on the point-set from

$$S = \frac{1}{N}\sum_{\vec{n}} \sigma_{\vec{n}}^2 |\sum_i e_{\vec{n}}(\vec{x})|^2 = \frac{1}{NM}\sum_{|\vec{n}|\le m} |\sum_i e_{\vec{n}}(\vec{x})|^2 \equiv \frac{1}{NM}\sum_{i,j}\psi(\vec{x}_i - \vec{x}_j) \tag{3.93}$$

with

$$\psi(\vec{x}_i - \vec{x}_j) = -1 + \prod_{\mu=1}^{D} \frac{\sin\Big((2m+1)\pi(x_i^\mu - x_j^\mu)\Big)}{\sin\Big(\pi(x_i^\mu - x_j^\mu)\Big)} \tag{3.94}$$

The distribution of point-sets with a particular value for s is then found by explicitly performing the z-integrals of eq.(3.72):

$$H(s) = \frac{K^K s^{K-1}}{\Gamma(K)} e^{-Ks} \tag{3.95}$$

with $K \equiv M/2$. Hence the correlation function is

$$F(s; \vec{x}_i - \vec{x}_j) = \frac{(1-s)}{M}\psi(\vec{x}_i - \vec{x}_j) \tag{3.96}$$

and the estimator[29] of eq.(3.80) becomes

$$E_2^{(q)} = \frac{1}{N^2}\sum f_i^2 - \frac{1}{N^3}(\sum f_i)^2 - \frac{1}{N^3}\frac{(1-s)}{M}\sum_{i,j}\psi(\vec{x}_i - \vec{x}_j) f_i f_j \tag{3.97}$$

This form has the advantage of including all modes up to an arbitrary m without much effort, with the overhead, of course, of being quadratic in N. As N grows beyond 10^5 this becomes particularly impractical. For investigating purposes, however, this approach is useful in testing the behavior of $E_2^{(q)}$ with more modes included (that is presumably the small λ limit).

[29]The use of the improved estimator of eq.3.85 in the box approximation is prohibited by the quadruple sums that it would contain.

It is remarkable that in the limit $m \to \infty$ we have $\psi(\vec{x}_i - \vec{x}_j) = M\delta_{i,j}$, and this leads to $s = 1$

$$E_2^{(q)} = \left(\frac{1}{N^2} \sum f_i^2 - \frac{1}{N^3} (\sum f_i)^2 \right) \tag{3.98}$$

In that limit a good point-set would have to integrates well any mode using a finite number of points N. Since that is impossible, all point-sets will be evaluated as equally bad by the particular diaphony.

It is evident that one has to find an optimal value for m. In the following plot the estimator $E_2^{(q)}$ is shown for TF5 in 2 dimensions with different values for m ranging from 3 to 30.

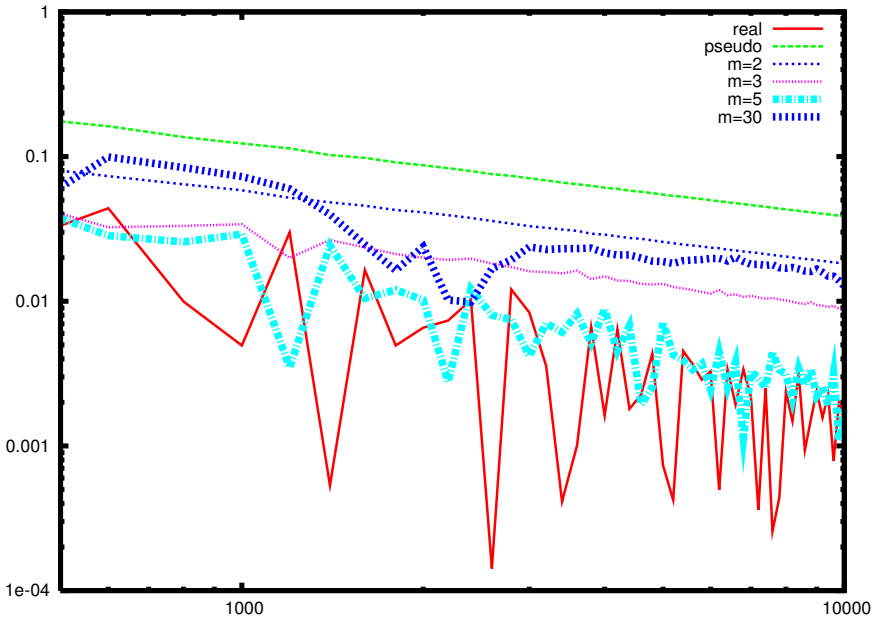


Figure 3.13: The box approximation: TF6, d=2 log-plot of the real error, and then from top to bottom the classical estimate, E_2^q with $m = 2$, with $m = 3$ and $m = 5$. The more modes one adds to the estimator the better it behaves. We also include the case $m = 30$ (thick dotted line), to demonstrate that there is a turning point in m above which the estimate becomes worse. Note that $m = 5$ means square length up to $2m^2 = 50$, much higher than 15 that was our ceiling in the plots of the previous sections.

3.7 Concluding remarks

Let us close this chapter by recapitulating the main conclusions of our analysis:

- The use of Quasi-Monte Carlo point-sets in numerical integration achieves a smaller error than the use of pseudo-random Monte Carlo point-sets. This advantage cannot be put in use without a reliable method for estimating the integration error.
- The 'classical', stochastic, error estimator relies on the assumption that the points in the point-set are uncorrelated. When used with a Quasi-Monte Carlo point-set, this assumption no longer holds. We saw that this leads to overestimating the error, thereby canceling any advantage gained by using the Quasi-Monte Carlo point-set.
- An estimator of stochastic nature is still possible but the underlying ensemble can not be the ensemble of all point-sets. We advocate the use of the ensemble of point-sets with the same degree of uniformity, as measured by a chosen diaphony. This approach leads to a prescription for a correlation function and an estimator, without the use of any information on the particular point-set or integrand.
- The price to pay is the raise in the computational complexity of the estimator from linear to quadratic in the number of points, which reflects the inclusion in the estimator of correlations between pairs of points. Using properties of diaphonies one can revert to a complexity that is linear times the number of modes involved.
- The error estimator suggested in this paper is shown to perform better than the 'classical' error estimator, resulting in an estimate up to an order of magnitude smaller than the 'classical' one.
- The flexibility of the construction (reflected in the freedom to choose the precise diaphony and the number of modes included) allows one to trade accuracy for computational cost. In computationally expensive applications, the monitoring approach of section 3.6.2 could be used to obtain an estimate that lies somewhere between the 'classical' and the quasi regime.

Chapter 4

BCFW recursive relations in Yang-Mills theories

A proof of the BCFW recursion relation for color-decomposed amplitudes with gluons is presented. The proof is based on a direct correspondence between terms in the BCFW decomposition and Feynman diagrams in a selected gauge.

4.1 Introduction to spinor techniques, MHV amplitudes and the BCFW decomposition

Spinors, the objects that appear in the spin-1/2 representation of the Lorentz group, and, therefore, in all Feynman diagrams containing fermions, are objects well-studied since the work of Cartan [17], with a multitude of useful properties. It should come as no surprise that they have been used repeatedly in the past in many approaches to calculating efficiently tree-level Feynman diagrams. Particularly in the high energy limit (i.e. when all fermion masses involved in the process we have in mind can safely be ignored), spinor techniques, developed since 1986 [41], achieved to calculate tree-level amplitudes for processes with many outgoing particles in a very compact form. These techniques used spinors to express polarization vectors of external gauge bosons, and took advantage of helicity conservation to end up in compact formulas for the matrix element itself, in terms of spinor products. This should be compared to the conventional method, whereby one has to evaluate the matrix element, then square it, sum over helicities and use trace identities to find the, much longer, final expression in terms of the external four-momenta. Spinor techniques for massive particles (fermions or bosons) were readily available.

In the same year, in a paper of exceptional brevity [53], Parke and Taylor conjectured a remarkably simple form for the tree-level color ordered amplitudes that involve only gluons and have a particular helicity structure [1]. These formulas where proven later by Berends and Giele [5] with the use of recursive relations for color-ordered amplitudes. The Parke-Taylor formula can be expressed in terms of spinor products alone. In the last couple of years there has been an impressive revival of interest on spinor techniques. A paper by E.Witten [69] made a connection between perturbative pure Yang-Mills theories and a certain topological string theory in twistor space. This triggered another paper by F.Cachazo, F.Svrcek and E.Witten [15], where a conjecture[2] is made that any helicity color-ordered amplitude can be calculated using effective vertices that are themselves continuations of Parke-Taylor amplitudes with some legs off-shell. This latter result was considered promising enough to be pursued in every other possible direction, including extensions to amplitudes involving fermions, particles with masses, particles from the electroweak sector [24], [4], [11], and a novel approach to one-loop amplitudes [12], [10] that also employs unitarity arguments. In all the above, Weyl-Van der Waerden spinors have been instrumental[3].

Moreover, Britto, Cachazo and Feng (BCFW) have proposed a new recursion relation for tree amplitudes of gluons [57] based on a different analytic continuation of Parke-Taylor-like objects, that naturally arrives at the simplest known expressions for some of those amplitudes in terms of Weyl - Van der Waerden spinor products. Explicit calculations have been performed using this technique [33], [32], [43], [20] and an extension to fermions has been conjectured [70], [71], [73], [52]. A proof of the recursive relation was immediately given by Britto, Chachazo, Feng and Witten [14] which, though it made the connection between the analyticity properties of the color amplitude and the BCFW decomposition obvious, shed little light on how the latter

[1] They also challenged string theorists to prove their result rigorously!

[2] The CSW decomposition in MHV vertices has been related with the BCFW recursion relation in [58] and derived from a Lagrangian after a canonical transformation related to the light-cone gauge in [47] and [30]. A direct proof in terms of Feynman diagrams is still missing.

[3] Despite the fact that one can always express everything in terms of Dirac spinors or four-momenta products, if so wished.

relates to the usual Feynman diagram approach.

The simplicity of the final result for tree-level gluonic amplitudes implies the existence of major cancelations between Feynman graphs. Since the BCFW formalism avoids those cancelations altogether it is advantageous to find a diagrammatic proof of the BCFW decomposition in order to understand these cancelations in the level of Feynman diagrams. A number of questions immediately arise. Is a diagrammatic proof gauge-dependent? If so, are there preferable gauges where simplifications take place? Which gauges are those? Is this decomposition exclusively working for pure YM theories, and if so, what is the key feature of the YM vertices that makes it work? Can we find a BCFW type recursion relation for scalar theories?

In what follows, after defining spinors and clarifying our notation, we will present shortly some background knowledge on color-decomposition and the Parke-Taylor formula, and then give a full diagrammatic correspondence between terms of the BCFW decomposition and the ordinary Feynman diagrams. It will hopefully become clear that the BCFW decomposition is a re-arrangement of Feynman diagram terms in a convenient way, and that no deeper principle needs to be evoked to explain the simplicity of its results.

4.2 Spin 1/2 particles

In this section we will define, beginning from the Dirac representation of the Lorentz group, the spinors that we will use subsequently, both in the Dirac and in Weyl - Van der Waerden form, as well as the spinorial representation of gauge bosons' polarization vectors, along with a rich, and usually confusing, formalism of dotted and undotted indices, spinor products and the like. The discussion is kept short, but hopefully clarifying, and the reader is deferred to the references for further details.

4.2.1 Dirac representation of the Lorentz group

All irreducible representations of the Lorentz group can be built as direct products of the tensor and the spinor representation. In what follows we will construct the spinor representation of the group.

Any representation of the Lorentz group has to satisfy

$$U(\Lambda)U(\overline{\Lambda}) = U(\Lambda\overline{\Lambda}) \tag{4.1}$$

One can find immediately the commutation relations of the Lorentz algebra by evaluating

$$U(\Lambda)U(\overline{\Lambda})U(\Lambda^{-1}) = U(\Lambda\overline{\Lambda}\Lambda^{-1}) \tag{4.2}$$

separately on the left and right hand side, in the case of infinitesimal transformations

$$U(\Lambda) = 1 + \frac{i}{2}\omega_{ab}J^{ab} \tag{4.3}$$

Equating coefficients of ω one finds

$$[J^{ab}, J^{cd}] = i(g^{ac}J^{bd} - g^{ad}J^{bc} + g^{bd}J^{ac} - g^{bc}J^{ad}) \tag{4.4}$$

In the spinor representation the above commutation relations are satisfied by

$$J^{ab} = \frac{-i}{4}[\gamma^a, \gamma^b] \tag{4.5}$$

with γ's objects that satisfy

$$\{\gamma^a, \gamma^b\} = 2g^{ab} \cdot \infty \tag{4.6}$$

In 4 dimensions a convenient representation of the *gamma*-matrices is the Weyl representation

$$\gamma^0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ \overline{\sigma^i} & 0 \end{pmatrix} \tag{4.7}$$

with $\overline{\sigma^i} \equiv -\sigma^i$, σ^i being the usual Pauli matrices

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{4.8}$$

We can also define γ^5

$$\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \tag{4.9}$$

with which we can form a complete set of matrices in the space of 4×4 matrices, by

$$1, \gamma^\mu, J^{\mu\nu}, \gamma^5\gamma^\mu, \gamma^5 \tag{4.10}$$

It is also useful to know that

$$\gamma^0\gamma^{\mu\dagger}\gamma^0 = \gamma^\mu \tag{4.11}$$

and, in the Weyl representation,

$$\gamma^{\mu T} = -C\gamma^\mu C \quad C = \gamma^2\gamma^0 \tag{4.12}$$

which is equivalent to

$$\overline{\sigma}^{\mu T} = \sigma^2\sigma^\mu\sigma^2 \qquad \sigma^{\mu T} = \sigma^2\overline{\sigma}^\mu\sigma^2 \tag{4.13}$$

4.2.2 Spinors for massless spin 1/2 particles

There are many ways to define the four-component Dirac spinors one is supposed to use in Feynman diagrams. One way is to follow Weinberg [67] and define the physical spinors as coefficients of the creation and annihilation operators of a physical spin 1/2 field. The transformation properties of the spinors under the Lorentz group are then sufficient to define them.

Another way is to define the spinors as the objects one should attach to fermion lines in Feynman diagrams that have been truncated. The completeness relation becomes then a defining equation relating the spinors with the truncated Dirac propagator. Considerations in Clifford algebra can then show that the minimum dimension of γ matrices is 4×4 and that the only projection operator possible within this algebra, for massive particles, will have the form

$$\Pi(k,s) = \frac{1}{4}(1 \pm \not{k})(1 \pm \gamma^5 \not{s}) \quad k^2 = 1 = -s^2 \tag{4.14}$$

where $s^2 = -1$ is to be interpreted as a spin vector and $K^2 = 1 = \frac{p^2}{m}$. The massless limit would be

$$\Pi(k,\pm) = \frac{1}{2}(1 \pm \gamma^5)\not{k} \quad k^2 = 0 \tag{4.15}$$

from which one can define four four-component Dirac-spinors for every external particle: $u_+(p)$, $u_-(p)$ for particles going into the amplitude and $\overline{u}_+(p)$, $\overline{u}_-(p)$ for particles going out of the amplitude.

Note that the two Dirac spinors are not necessarily connected by complex conjugation or the commonly seen bar operation (complex conjugation and right multiplication with γ^0). Moreover $\overline{u}_h(p)$ can be perceived as a matrix with one row and four columns (a row spinor as opposed to the column spinor $u_h(p)$). Here h is the helicity-related degree of freedom.

The defining equations for the spinors are

$$\frac{1}{2}(1+\gamma_5)\not{p} = u_+(p)\overline{u}_+(p) \quad \frac{1}{2}(1-\gamma_5)\not{p} = u_-(p)\overline{u}_-(p) \tag{4.16}$$

and

$$\begin{aligned} \frac{1}{2}(1+\gamma_5)u_+(p) &= u_+(p) \qquad & \frac{1}{2}(1-\gamma_5)u_-(p) &= u_-(p) \\ \overline{u}_+(p)\frac{1}{2}(1-\gamma_5) &= \overline{u}_+(p) \qquad & \overline{u}_-(p)\frac{1}{2}(1+\gamma_5) &= \overline{u}_-(p) \end{aligned} \tag{4.17}$$

In The Weyl representation

$$\frac{1}{2}(1+\gamma^5) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \quad \frac{1}{2}(1-\gamma^5) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \tag{4.18}$$

which, in view of eq.(4.17) forces two of the four components of each u to vanish. We have

$$u_+(p) = \begin{pmatrix} 0 \\ \phi \end{pmatrix} \quad u_-(p) = \begin{pmatrix} \chi \\ 0 \end{pmatrix} \tag{4.19}$$

$$\overline{u}_+(p) = \begin{pmatrix} \tilde{\phi} & 0 \end{pmatrix} \quad \overline{u}_-(p) = \begin{pmatrix} 0 & \tilde{\chi} \end{pmatrix} \tag{4.20}$$

where ϕ,χ,$\tilde{\phi}$ and $\tilde{\chi}$ are two component spinors.

The defining equation, eq.(4.16), then turns into a condition for these two component "Weyl-Van der Waerden" spinors:

$$p \cdot \sigma = \chi\tilde{\chi} \quad p \cdot \overline{\sigma} = \phi\tilde{\phi} \tag{4.21}$$

Let us define

$$\epsilon \equiv \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{4.22}$$

Thanks to the identity

$$\overline{\sigma}^{\mu} = -\epsilon(\sigma^{\mu})^{T}\epsilon \tag{4.23}$$

which is equivalent with eq.(4.13), we can write

$$\phi\tilde{\phi} = p\cdot\overline{\sigma} = -\epsilon p_{\mu}(\sigma^{\mu})^{T}\epsilon = -\epsilon(\chi\tilde{\chi})^{T}\epsilon = -\epsilon\tilde{\chi}^{T}\chi^{T}\epsilon \tag{4.24}$$

This implies a relation between ϕ's and χ's up to (once again) a scale factor. We can (even though it's not obligatory) define

$$\phi = -\epsilon\tilde{\chi}^{T} \quad \tilde{\phi} = \chi^{T}\epsilon \tag{4.25}$$

4.2.3 Dotted and undotted indices

At this point we will introduce the notation with the dotted and undotted indices which is overwhelmingly popular in the supersymmetry and superstrings literature. We denote

$$p_{a\dot{a}} \equiv p_{\mu}\sigma^{\mu}_{a\dot{a}} \tag{4.26}$$

If, and only if, $p^2 = 0$, the matrix $p_{a\dot{a}}$ can be written as a tensor product of two spinors, denoted by $\lambda_a(p), \tilde{\lambda}_{\dot{a}}(p)$

$$p^2 = 0 \Rightarrow p_{a\dot{a}} = \lambda_a(p)\tilde{\lambda}_{\dot{a}}(p) \tag{4.27}$$

Let us introduce the antisymmetric matrix

$$\epsilon^{ab} = \epsilon^{\dot{a}\dot{b}} \equiv \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{4.28}$$

In index notation, eq.(4.23) becomes

$$\overline{\sigma}^{\mu,\dot{a}a} = -\epsilon^{\dot{a}\dot{b}}(\sigma^{\mu\,T})_{\dot{b}c}\epsilon^{ca} \tag{4.29}$$

which leads to the very useful identity

$$2g^{\mu\nu} = Tr\{\sigma^{\mu}\overline{\sigma}^{\nu}\} = \sigma^{\mu}_{a\dot{a}}\overline{\sigma}^{\nu,\dot{a}a} = \sigma^{\mu}_{a\dot{a}}\sigma^{\nu}_{b\dot{b}}\epsilon^{ab}\epsilon^{\dot{a}\dot{b}} \tag{4.30}$$

We can, then, define the scalar products

$$\langle pq\rangle \equiv \lambda_a(p)\lambda_b(q)\epsilon^{ab} \tag{4.31}$$

$$[pq] \equiv \tilde{\lambda}_{\dot{a}}(p)\tilde{\lambda}(q)_{\dot{b}}\epsilon^{\dot{a}\dot{b}} \tag{4.32}$$

and note that thanks to eq.(4.30)

$$\langle pq\rangle[pq] = \lambda_a(p)\lambda_b(q)\epsilon^{ab}\tilde{\lambda}_{\dot{a}}(p)\tilde{\lambda}(q)_{\dot{b}}\epsilon^{\dot{a}\dot{b}} = p_{\mu}q_{\nu}\sigma^{\mu}_{a\dot{a}}\sigma^{\nu}_{b\dot{b}}\epsilon^{ab}\epsilon^{\dot{a}\dot{b}} = p_{\mu}q_{\nu}2g^{\mu\nu} = 2p\cdot q \tag{4.33}$$

In general, for any two four-vectors p^{μ}, q^{μ}, we have

$$p\cdot q = p_{\mu}q_{\nu}g^{\mu\nu} = \frac{1}{2}p_{\mu}q_{\nu}\sigma^{\mu}_{a\dot{a}}\sigma^{\nu}_{b\dot{b}}\epsilon^{ab}\epsilon^{\dot{a}\dot{b}} = \frac{1}{2}p_{a\dot{a}}q_{b\dot{b}}\epsilon^{ab}\epsilon^{\dot{a}\dot{b}} \tag{4.34}$$

We proceed further and define spinors with upper indices, using ϵ to raise indices as follows

$$\lambda^{a} = \lambda_{b}\epsilon^{ba} \qquad \tilde{\lambda}^{\dot{a}} = \epsilon^{\dot{a}\dot{b}}\tilde{\lambda}_{\dot{b}} \tag{4.35}$$

Note that undotted indices are raised by right multiplication whereas dotted ones by left multiplication.

Let us define $\epsilon_{ab}, \epsilon_{\dot{a}\dot{b}}$, by

$$\epsilon_{ab}\epsilon^{bc} = -\delta_a^c \Longleftrightarrow \epsilon_{ab}\epsilon^{cb} = \delta_a^c \tag{4.36}$$

$$\epsilon_{\dot{a}\dot{b}}\epsilon^{\dot{b}\dot{c}} = -\delta_{\dot{a}}^{\dot{c}} \Longleftrightarrow \epsilon_{\dot{a}\dot{b}}\epsilon^{\dot{c}\dot{b}} = \delta_{\dot{a}}^{\dot{c}} \tag{4.37}$$

which gives

$$\epsilon_{ab} = \epsilon_{\dot{a}\dot{b}} \equiv \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \epsilon^{ab} = \epsilon^{\dot{a}\dot{b}} \tag{4.38}$$

Lowering indices can then be performed by multiplying eq.(4.35) with the appropriate lower index ϵ

$$\epsilon_{ca}\lambda^a = \epsilon_{ca}\lambda_b\epsilon^{ba} = \lambda_c \tag{4.39}$$

$$\tilde{\lambda}^{\dot{a}}\epsilon_{\dot{a}\dot{c}} = \epsilon^{\dot{a}\dot{b}}\tilde{\lambda}_{\dot{b}}\epsilon_{\dot{a}\dot{c}} = \tilde{\lambda}_{\dot{c}} \tag{4.40}$$

At this point it is worth introducing the identity

$$p_a q_b = q_a p_b + \epsilon_{ab}\langle pq\rangle \tag{4.41}$$

and the corresponding one for square bracket products

$$\tilde{p}_{\dot{a}}\tilde{q}_{\dot{b}} = \tilde{q}_{\dot{a}}\tilde{q}p_{\dot{b}} + \epsilon_{\dot{a}\dot{b}}[pq] \tag{4.42}$$

Both of them can easily be proven by multiplying with ϵ^{ab} and $\epsilon\dot{a}\dot{b}$ respectively. Using them one can prove the Schouten identity

$$\langle pq\rangle s_c = p_a q_b s_c \epsilon^{ab} = p_a s_b q_c \epsilon^{ab} + p_a\langle qs\rangle\epsilon_{bc}\epsilon^{ab} = -\langle qs\rangle p_c - \langle sp\rangle q_c \tag{4.43}$$

and similarly for the square brackets

$$[pq]\tilde{s}_{\dot{c}} = -[qs]\tilde{p}_{\dot{c}} - [sp]\tilde{q}_{\dot{c}} \tag{4.44}$$

The connecting formula with the ordinary Dirac-spinor notation of the previous section is provided by eq.(4.27):

$$p\cdot\sigma_{a\dot{a}} = \lambda_a(p)\tilde{\lambda}_{\dot{a}}(p) = \chi\tilde{\chi} \tag{4.45}$$

from which

$$\chi \equiv \lambda_a \quad \tilde{\chi} \equiv \tilde{\lambda}_{\dot{a}} \tag{4.46}$$

Furthermore,

$$p\cdot\overline{\sigma}^{\dot{a}a} = p_\mu\overline{\sigma}^{\mu,\dot{a}a} = -p_\mu\epsilon^{\dot{a}\dot{b}}(\sigma^{\mu\,T})_{\dot{b}c}\epsilon^{ca} = -\epsilon^{\dot{a}\dot{b}}\tilde{\lambda}_{\dot{b}}\lambda_c\epsilon^{ca} = -\tilde{\lambda}^{\dot{a}}\lambda^a = \phi\tilde{\phi} \tag{4.47}$$

from which

$$\phi = -\tilde{\lambda}^{\dot{a}} \quad \tilde{\phi} = \lambda^a \tag{4.48}$$

To summarize, the Dirac spinors with definite helicity can be written as

$$u_+(p) = \begin{pmatrix} 0 \\ -\tilde{\lambda}^{\dot{a}} \end{pmatrix} \quad u_-(p) = \begin{pmatrix} \lambda_a \\ 0 \end{pmatrix} \tag{4.49}$$

$$\overline{u}_+(p) = \begin{pmatrix} \lambda^a & 0 \end{pmatrix} \quad \overline{u}_-(p) = \begin{pmatrix} 0 & \tilde{\lambda}_{\dot{a}} \end{pmatrix} \tag{4.50}$$

It is worth mentioning here that all the spinor definitions of the present section refer to spinors that correspond to on-shell external particles. Spinors corresponding to momenta that are not on-shell have to be defined in some way, as we shall see, but the above definitions are certainly not valid in that case.

Clarifying this, we should note the general result that any 2×2 complex matrix with zero determinant can be written as a dyad of two complex 2−spinors

$$det(X) = 0 \Rightarrow X_{a\dot{a}} = k_a \tilde{k}_{\dot{a}} \tag{4.51}$$

for some $k_a, \tilde{k}_{\dot{a}}$. This is the case for a massless momentum in the matrix representation $p \cdot \sigma$, since the vanishing of the determinant is easily seen to be equivalent with the fact that $p^2 = 0$. On the contrary, a matrix with non-zero determinant cannot be written as a dyad, but only as a sum of two dyads

$$det(X) \neq 0 \Rightarrow X_{a\dot{a}} = k_a \tilde{k}_{\dot{a}} + \mu_a \tilde{\mu}_{\dot{a}} \tag{4.52}$$

and hence any off-shell momentum, corresponding to a matrix with non-zero determinant, has no spinors directly connected with it, in the above mentioned way.

Due to their defining equation, eq.(4.27), the spinors λ_a and $\lambda_{\dot{a}}$ are only defined up to a scaling factor. By this we mean that one is free to perform the transformation

$$\lambda_a \to c\lambda_a \qquad \lambda_{\dot{a}} \to \frac{1}{c}\lambda_{\dot{a}} \tag{4.53}$$

without affecting the value of any measurable quantity that can be calculated with the help of these spinors. The number c is in principle complex. Due to a fundamental geometric property of spinors, it is actually not possible to define a scale factor c which would hold for all four-vectors p^μ in such a manner that the scale factor varies continuously with the four-vector[4]. In case the $\tilde{\lambda}_{\dot{a}} = \lambda_a^*$ constraint is imposed, the scale c has to be on the unit circle $c = e^{i\phi}$.

A final word on notation is due here. In what follows, we will often substitute the symbol λ with the symbol for the momentum to which it corresponds:

$$p_a \equiv \lambda_a(p) \qquad \tilde{p}_{\dot{a}} \equiv \tilde{\lambda}_{\dot{a}}(p) \tag{4.54}$$

Hence, we will be writing

$$p \cdot \sigma = p_a \tilde{p}_{\dot{a}} \quad p \cdot \overline{\sigma} = -\tilde{p}^{\dot{a}} p^a \tag{4.55}$$

Actually, in what follows we will sometimes even substitute the symbol p_i which denotes the momentum of the i'th particle, with the symbol i itself, writing equations like

$$p_3 \cdot \sigma = 3_a \tilde{3}_{\dot{a}} \quad p_3 \cdot \overline{\sigma} = -\tilde{3}^{\dot{a}} 3^a \tag{4.56}$$

4.2.4 Polarization vectors for massless spin-1 particles

Polarization vectors for gluons (or photons) need satisfy

$$\epsilon_\mu(k) \cdot k = 0 \;\; \epsilon^+ \cdot \epsilon^{+*} = -1 \;\; \epsilon^+ \cdot \epsilon^+ = 0 \tag{4.57}$$

as well as

$$\epsilon_\mu^+(k)\epsilon_\nu^{+*}(k) + \epsilon_\mu^-(k)\epsilon_\nu^{-*}(k) = D_{\mu\nu} \tag{4.58}$$

where $D_{\mu\nu} = S_{\mu\nu}k^2$ with $S_{\mu\nu}$ the gluon propagator in the selected gauge.

The fundamental idea behind all spinor techniques has been to express these polarization vectors in terms of the spinors that correspond to the four-momentum carried by each external gluon. In doing so there is a freedom of normalization and a freedom of phase, both of which are

[4] By this we mean that when one performs a rotation by 2π around, say, the z-axis for a given four-momentum, p^μ, the corresponding spinor is picking up a minus sign, whereas p^μ returns to itself (see [17] for details).

dropping out of every final result. One definition for these polarization vectors[5] is (see [22], [41] for similar definitions.)

$$\epsilon_+^\mu(p) = \frac{\overline{u}_+(q)\gamma^\mu u_+(p)}{\sqrt{2}\overline{u}_+(q)u_-(p)} = -\frac{\tilde{p}_{\dot{a}}\overline{\sigma}^{\mu,\dot{a}a}q_a}{\sqrt{2}\langle qp\rangle}$$
$$\epsilon_-^\mu(p) = \frac{\overline{u}_-(q)\gamma^\mu u_-(p)}{\sqrt{2}\overline{u}_-(p)u_+(q)} = -\frac{\tilde{q}_{\dot{a}}\overline{\sigma}^{\mu,\dot{a}a}p_a}{\sqrt{2}[pq]} \tag{4.59}$$

where q is an auxiliary, null four-vector that can be chosen at will, and independently for every gluon, as long as it is not parallel to p itself.

The equivalent in Weyl Van der Waerden notation is found by contracting the polarization vectors with $\sigma_{a\dot{a}}$ and using the identities [6]

$$\sigma^\mu_{a\dot{a}}\sigma_{\mu,b\dot{b}} = 2\epsilon_{ab}\tilde{\epsilon_{\dot{a}\dot{b}}} \quad \sigma^\mu_{a\dot{a}}\overline{\sigma}^{\dot{b}b}_\mu = 2\delta^b_a\delta^{\dot{b}}_{\dot{a}} \tag{4.60}$$

We find

$$\epsilon^+_{a\dot{a}}(p) \equiv \epsilon^+_\mu(p)\sigma^\mu_{a\dot{a}} = \sqrt{2}\frac{q_a\tilde{p}_{\dot{a}}}{\langle qp\rangle} \tag{4.61}$$

$$\epsilon^-_{a\dot{a}}(p) \equiv \epsilon^-_\mu(p)\sigma^\mu_{a\dot{a}} = -\sqrt{2}\frac{p_a\tilde{q}_{\dot{a}}}{[pq]} \tag{4.62}$$

As a check, the dot product of the two vectors is

$$\epsilon^+\cdot\epsilon^- = \frac{1}{2}\epsilon^+_{a\dot{a}}\epsilon^-_{b\dot{b}}\epsilon^{ab}\epsilon^{\dot{a}\dot{b}} = -\frac{1}{2}\sqrt{2}\frac{q_a\tilde{p}_{\dot{a}}}{\langle qp\rangle}\sqrt{2}\frac{p_a\tilde{q}_{\dot{a}}}{[pq]}\epsilon^{ab}\epsilon^{\dot{a}\dot{b}} = -\frac{\langle qp\rangle[pq]}{\langle qp\rangle[pq]} = -1 \tag{4.63}$$

and the tensor product is

$$D_{\mu\nu}\sigma^\mu_{a\dot{a}}\sigma^\nu_{b\dot{b}} = \{\epsilon^+_\mu(k)\epsilon^-_\nu(k) + \epsilon^-_\mu(k)\epsilon^+_\nu(k)\}\sigma^\mu_{a\dot{a}}\sigma^\nu_{b\dot{b}} \tag{4.64}$$
$$= \epsilon^+_{a\dot{a}}\epsilon^-_{b\dot{b}} + \epsilon^-_{a\dot{a}}\epsilon^+_{b\dot{b}} = \tag{4.65}$$
$$= -2\frac{q_a\tilde{p}_{\dot{a}}p_b\tilde{q}_{\dot{b}} + q_b\tilde{p}_{\dot{b}}p_a\tilde{q}_{\dot{a}}}{\langle qp\rangle[pq]} \tag{4.66}$$

We can use the identities eq.(4.41), eq.(4.42) to match the terms of the numerator in eq.(4.66), so that

$$D_{\mu\nu}\sigma^\mu_{a\dot{a}}\sigma^\nu_{b\dot{b}} = -2\frac{p\cdot\sigma_{a\dot{a}}q\cdot\sigma_{b\dot{b}} + q\cdot\sigma_{a\dot{a}}p\cdot\sigma_{b\dot{b}}}{\langle qp\rangle[pq]} - \epsilon_{ab}\epsilon_{\dot{a}\dot{b}} = \tag{4.67}$$
$$= \{-g_{\mu\nu} + \frac{p_\mu q_\nu + p_\nu q_\mu}{p\cdot q}\}\sigma^\mu_{a\dot{a}}\sigma^\nu_{b\dot{b}} \tag{4.68}$$

which, finally, shows that the above choice of polarization vectors is equivalent to the axial gauge with q in the role of the axial vector. It should be noted here that a change in the auxiliary momentum q, resulting at a change in the spinors $\lambda_a(q), \tilde{\lambda}_{\dot{a}}(q)$, is fully equivalent to a change of gauge for the particular gluon. In other words, a transformation

$$\epsilon^\mu_+ \to \epsilon^\mu_+ + cp^\mu \tag{4.69}$$

would result to

$$\epsilon^+_{a\dot{a}}(p) \to \epsilon^+_{a\dot{a}}(p) + cp_a\tilde{p}_{\dot{a}} = \sqrt{2}\frac{p_a\tilde{q}'_{\dot{a}}}{\langle p'p\rangle} \tag{4.70}$$

[5] That corresponds to the light-like axial gauge.

[6] These identities are equivalent to the Chisholm identity and related to the Fierz rearrangement lemma.

with

$$q'_a = q_a + \frac{c}{\sqrt{(2)}}\langle qp\rangle p_a \tag{4.71}$$

Conversely, one can express any choice of q as a change of ϵ in the sense of eq.(4.69). We will make abundant use of our freedom, in view of the Ward identities, to pick auxiliary vectors separately for every external gluon. Judicious choice of the auxiliary vectors makes various spinor products vanish and hence various Feynman graphs cancel. The most usual choice has been (see [22]) to use the four-momentum of one '+'-helicity leg as auxiliary vector for all the '−'-helicity legs and vice versa.

4.3 Color decomposition - The Feynman rules for dual amplitudes in QCD

In this section we review shortly the trace-based color-decomposition technique that results to the so-called color-ordered amplitudes. It has long been observed ([46], [7]) that one can take advantage of group theory identities in order to decompose any given amplitude, in massless QCD, into a sum of color-ordered or dual amplitudes, i.e. amplitudes whose color structure appears as a single multiplicative factor. These amplitudes consist of color-ordered Feynman diagrams, i.e. diagrams where the order in which the external legs appear is fixed.

The Feynman rule for the three-gluon vertex involves a color factor

$$f^{abc} = -2iTr(t^a t^b t^c - t^a t^c t^b) \tag{4.72}$$

whereas the one of the four-gluon vertex involves three terms with color factors of the form $f^{kab}f^{kcd}$ (summation over k is as always implied). The latter can be transformed into traces with the help of another group-theory identity

$$(t^a)_{ij}(t^a)_{km} = \frac{1}{2}\delta_{im}\delta_{jk} - \frac{1}{2N_c}\delta_{ij}\delta_{km} \tag{4.73}$$

because

$$f^{kab}f^{kcd} = -4Tr(t^k t^a t^b - t^k t^b t^a)Tr(t^k t^c t^d - t^k t^d t^c) \tag{4.74}$$

and

$$\begin{aligned} Tr(t^k t^a t^b)Tr(t^k t^c t^d) &= (t^k_{ij}t^a_{jk}t^b_{ki})(t^k_{mn}t^c_{nr}t^d_{rm}) = & (4.75)\\ &= \frac{1}{2}t^a_{jk}t^b_{ki}t^c_{ir}t^d_{rj} - \frac{1}{2N_c}t^a_{jk}t^b_{kj}t^c_{ir}t^d_{ri} = & (4.76)\\ &= \frac{1}{2}Tr(t^a t^b t^c t^d) - \frac{1}{2N_c}Tr(t^a t^b)Tr(t^c t^d) & (4.77)\end{aligned}$$

Denoting

$$Tr(t^a t^b \dots t^c) \equiv (ab\dots c) \tag{4.78}$$

we, then, have for the four-vertex factor

$$f^{kab}f^{kcd} = -2\left[(abcd) - (acbd) - (bacd) + (badc)\right] \tag{4.79}$$

This means that the four-gluon vertex can be written as

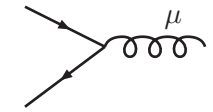
$$\begin{aligned} & -ig^2[f^{kac}f^{kbd}(g_{\mu\nu}g_{\rho\sigma} - g_{\mu\sigma}g_{\nu\rho}) + f^{kad}f^{kbc}(g_{\mu\nu}g_{\rho\sigma} - g_{\mu\rho}g_{\nu\sigma}) + f^{kab}f^{kcd}(g_{\mu\rho}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\rho})]\\ = \;& 2ig^2[((acbd) - (cabd) - (acdb) + (cadb))(g_{\mu\nu}g_{\rho\sigma} - g_{\mu\sigma}g_{\nu\rho})\\ & +((adbc) - (dabc) - (adcb) + (dacb))(g_{\mu\nu}g_{\rho\sigma} - g_{\mu\rho}g_{\nu\sigma})\\ & +((abcd) - (bacd) - (abdc) + (badc))(g_{\mu\rho}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\rho})]\\ = \;& 2ig^2[(acbd)(g_{\mu\nu}g_{\rho\sigma} - g_{\mu\sigma}g_{\nu\rho} + g_{\mu\nu}g_{\rho\sigma} - g_{\mu\rho}g_{\nu\sigma})\\ & +(abdc)(-g_{\mu\nu}g_{\rho\sigma} + g_{\mu\sigma}g_{\nu\rho} - g_{\mu\rho}g_{\nu\sigma} + g_{\mu\sigma}g_{\nu\rho})\\ & +(acdb)(-g_{\mu\nu}g_{\rho\sigma} + g_{\mu\sigma}g_{\nu\rho} - g_{\mu\rho}g_{\nu\sigma} + g_{\mu\sigma}g_{\nu\rho})\\ & +(adbc)(g_{\mu\nu}g_{\rho\sigma} - g_{\mu\sigma}g_{\nu\rho} + g_{\mu\nu}g_{\rho\sigma} - g_{\mu\rho}g_{\nu\sigma})\\ & +(abcd)(-g_{\mu\nu}g_{\rho\sigma} + g_{\mu\rho}g_{\nu\sigma} + g_{\mu\rho}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\rho})\\ & +(adcb)(-g_{\mu\nu}g_{\rho\sigma} + g_{\mu\rho}g_{\nu\sigma} + g_{\mu\rho}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\rho})] \end{aligned} \tag{4.80}$$

Hence each four-vertex can be analyzed in six pieces in each of which the color information is encoded as a multiplicative factor in front of the space-time part. The six pieces represent

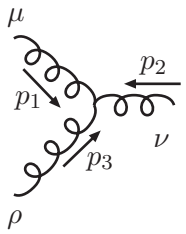
the six different ways in which one can order cyclically four objects. Note that the space-time structure of each piece is completely determined by the ordering: if the ordering is $(abcd)$ with corresponding Lorentz indices $\mu\nu\rho\sigma$, the space-time part contains a 'cross-term' with a factor of 2, $2g_{\mu\rho}g_{\nu\sigma}$, and two 'planar' terms $-g_{\mu\nu}g_{\rho\sigma} - g_{\mu\sigma}g\nu\rho$. Thanks to this property one can decompose every Feynman diagram into a number of color-ordered diagrams, one for each possible ordering of the external legs.

The Feynman rules for the color-ordered graphs are

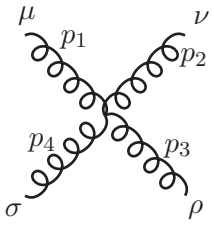
Table 4.1: QCD Feynman rules for color-ordered graphs



$-ig\gamma^\mu$



$g[g_{\mu\nu}(p_1 - p_2)_\rho + g_{\nu\rho}(p_2 - p_3)_\mu + g_{\rho\mu}(p_3 - p_1)_\nu]$



$-ig^2(2g_{\mu\rho}g_{\nu\sigma} - g_{\mu\nu}g_{\rho\sigma} - g_{\mu\sigma}g_{\nu\rho})$

4.4 Parke-Taylor amplitudes

As seen in the previous section, one can calculate a helicity amplitude involving only external gluons, in tree-level, by calculating instead all the color-ordered helicity amplitudes and then summing them up with the proper color factors. The central issue in this approach is, therefore, the calculation of the n-gluon color-ordered amplitude with fixed helicities. To be precise, in order to be able to talk about helicities we have to define the direction of momenta of the external gluons. To avoid confusion we will always treat external momenta as in-going in what follows.

It can be shown (see [22] and [19] for the original remark) using spinor techniques, that if all the helicities of the external gluons are the same (all '+' or all '−') then the color-ordered amplitude vanishes (and so does the full helicity amplitude of course). The same is true when all but one helicities are the same. The first non-vanishing amplitudes come from helicities that are all but two the same. We call these 'Maximal Helicity Violating' amplitudes[7]. Let's call 'mostly plus' an amplitude in which most of the helicities are '+' and 'mostly minus' an amplitude in which most of the helicities are '−' Amplitudes with an even number of gluons and equal number of '+' and '−' fit in both the above classes, so they are mostly plus as well as mostly minus.

So the first non-vanishing mostly-plus amplitude, the mostly-plus MHV amplitude, is one with all helicities + except two which are −. Let us call the two gluons with negative helicity i, j. Since we deal with color-ordered amplitudes, the cyclic order in which the gluons appear in such an amplitude is fixed. The Parke-Taylor formula[8] for this amplitude [53] is then

$$A_{MHV+}(1^+, \ldots, i^-, \ldots, j^-, \ldots, n^+) = \frac{\langle ij \rangle^4}{\langle 12 \rangle \langle 23 \rangle \ldots \langle n-1n \rangle \langle n1 \rangle} \tag{4.81}$$

where the spinor products are defined in eq.4.32.

Similarly, the mostly minus amplitude can be immediately calculated by an equivalent formula with the angle brackets substituted by square brackets.

$$A_{MHV-}(1^-, \ldots, i^+, \ldots, j^+, \ldots, n^-) = \frac{[ij]^4}{[12][23] \ldots [n-1n][n1]} \tag{4.82}$$

where now we denoted the + legs with i, j. One remarkable characteristic of these formulas is that they only involve products of undotted (dotted) spinors, a property which has been described by the word homogeneity, reflecting the fact that the spinors corresponding to an on-shell momentum are defined up to a scaling phase, and the Parke-Taylor amplitudes are homogeneous of a given degree to a change of such a phase. The same property examined from the viewpoint of complex analysis, gives rise to the adjective 'holomorphic', stressing the fact that the amplitude is a function only of λ's (or only of $\tilde{\lambda}$'s) and not of their conjugates [9]. The holomorphicity of the MHV amplitudes has interesting geometric consequences in complex twistor space.

[7] The name has arisen presumably from the observation that an amplitude with all helicities + and all momenta in-going will look like $++--\ldots-$ when the momenta of the final state are flipped to be out-going as usual. Then helicity conservation is violated in the worse possible way. Of course, helicity needs not be conserved in color-ordered amplitudes.

[8] To be precise, in the original paper Parke and Taylor give the square of the formula below, which involves four-momenta products instead of spinor products, and describes the full helicity amplitude squared. The spinor-product form below is, nevertheless, widely referred to as the Parke-Taylor formula.

[9] In some generic definition of conjugation that needs not be restricted to complex conjugation.

4.5 BCFW decomposition in terms of Feynman diagrams

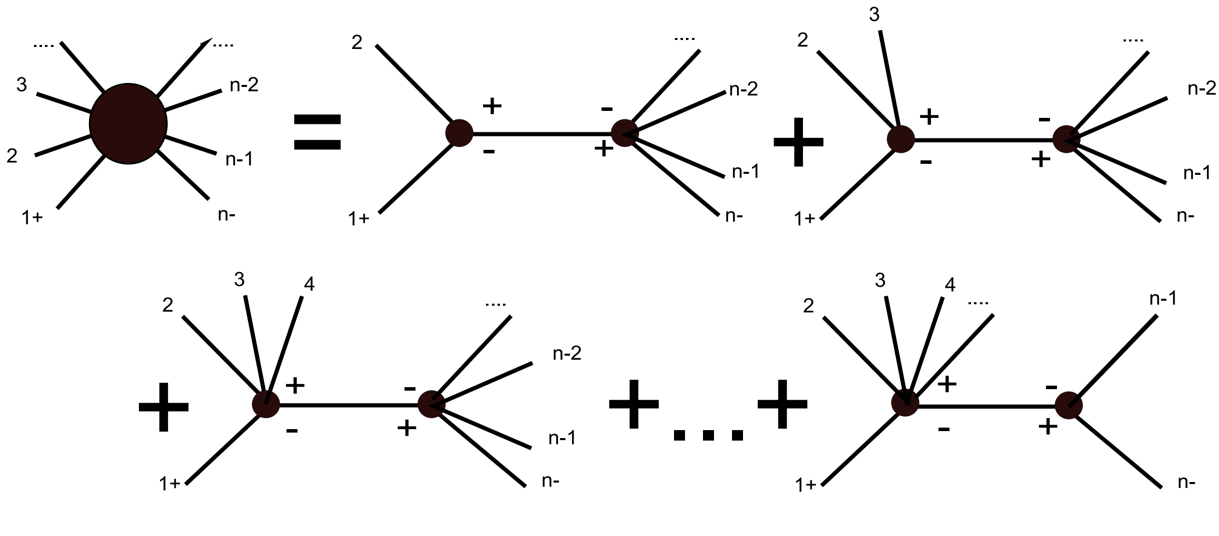
The BCFW recursion relation features some remarkable characteristics, among which the on-shell analytic continuation of selected off-shell propagators, the analytic continuation of two selected external momenta in the complex plane and a decomposition of a color helicity amplitude into smaller helicity amplitudes with complex external momenta that doesn't appear to be in direct connection with the decomposition in Feynman diagrams. Moreover, the BCFW relation leads to the use of only three-point (modified) vertices as building blocks of the theory, thus raising questions about the fundamental nature of the Yang-Mills four point vertex.

4.5.1 The BCFW recursion relation

Any color-ordered amplitude with n external legs is a function of the momenta p_i $(i = 1, \ldots, n)$ and the helicities h_i of those external legs. The familiar way to calculate such an amplitude would be to draw all possible Feynman diagrams with the particular ordering of the external legs and then calculate the corresponding expression for every diagram.

The BCFW decomposition proceeds in a different way. At first, one can always pick out a pair of consecutive[10] legs with opposite helicity (or else all legs have the same helicity and the color amplitude vanishes). Let's call the positive helicity leg by 1 and the negative one by n The color amplitude decomposes then to a sum of 'BCFW' graphs (see fig. 4.83).

The number of graphs is equal to $n - 3$, the number of ways the other $n - 2$ legs can be arranged in two groups preserving the cyclic ordering. Each graph is a product of two 'BCFW' vertices connected by a single propagator. The first and the last leg can never be in the same vertex.



(4.83)

Every 'BCFW' vertex with n legs corresponds to the expression for the full, on-shell, color amplitude with those n legs, with a modification in the momenta that we will describe shortly. The propagator connecting the two vertices is the scalar expression for the Feynman propagator, i.e. it amounts to $\frac{1}{Q^2}$ where Q^μ is the four-momentum carried in the internal line (which equals the sum of the external momenta of the left - or the right - vertex).

[10]In their original publication [57] it is shown that one can pick any two legs with opposite helicity, not necessarily consecutive. This is increasing the number of the BCFW decompositions needed without offering any advantage. To our knowledge, no application of the general decomposition has appeared since.

More precisely, the color amplitude $A(p_1^{h_1}, \ldots, p_n^{h_n})$ equals

$$A(p_1^{h_1}, \ldots, p_n^{h_n}) = \sum_{j=2}^{n-1} \sum_h \left[A(\hat{p}_1^{h_1}, \ldots, p_j^{h_j}, -\hat{P}_{1\ldots j}^h)\right]_{z=z_j} \frac{1}{P_{1\ldots j}^2} \left[A(\hat{P}_{1\ldots j}^{-h}, p_{j+1}^{h+1}, \ldots, \hat{p}_n^{h_n})\right]_{z=z_j} \tag{4.84}$$

where the sum over j extends over all partitions of the $n-2$ gluons in two groups. In the above expression, the hat symbol over a four-momentum denotes a particular analytic continuation of the denoted momenta in the complex plane, defined by

$$\begin{aligned} \hat{p}^\mu &= p^\mu + z\epsilon^\mu \quad p \neq p_n \\ \hat{p}^\mu &= p^\mu - z\epsilon^\mu \quad p = p_n \end{aligned} \tag{4.85}$$

with

$$\epsilon \cdot \sigma_{a\dot{a}} \equiv \epsilon_{a\dot{a}} = n_a \tilde{1}_{\dot{a}} \tag{4.86}$$

In eq.(4.84) the value of z_j is such that the internal leg $\hat{P} = p_1 + \ldots + p_j + z\epsilon$ is on-shell:

$$\hat{P}^2 = (p_1 + \ldots + p_j + z_i\epsilon)^2 = 0 \tag{4.87}$$

It is easy to see that the described recipe amounts to adding to the momentum of the first leg a complex four-vector $z\epsilon^\mu$, and subtracting it from the momentum of the last leg. Insisting on momentum conservation in every vertex, one finds that this complex four-vector has to be carried through by the connecting propagator. This complex four-vector is chosen such that it keeps the first and last legs on-shell, while the remaining freedom in z is used to set the internal legs on-shell as well!

The amplitude is recovered from the sum of $n-2$ decompositions in lower-level amplitudes evaluated using all the occurring $z = z_j$.

The above decomposition is recursive, in the sense that the smaller amplitudes, being amplitudes of (complex) on-shell momenta can be, in turn, decomposed further into even smaller amplitudes, and so forth, until one reaches the fundamental three-point Park-Taylor expression. In practice, of course, it is preferable to use the Parke-Taylor expression whenever an MHV amplitude appears in the decomposition process.

In the next section we present some examples to familiarize the reader with this rather odd way of calculating color-amplitudes.

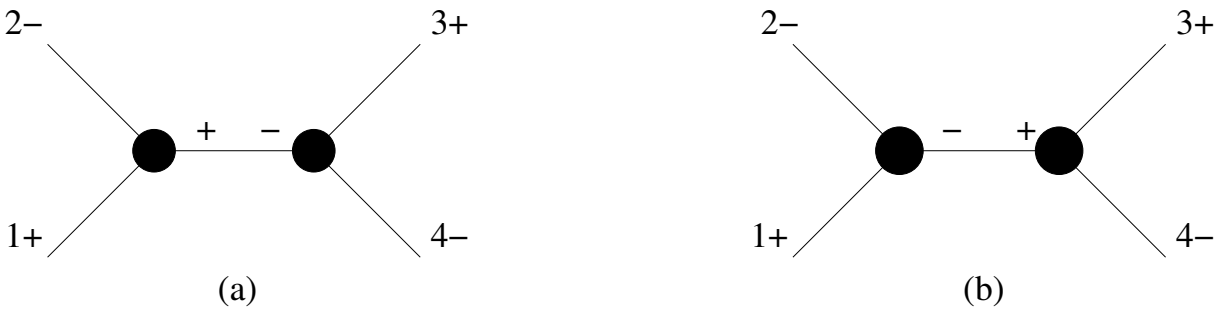
4.5.2 Examples

The four-point amplitude

As a first, trivial example, we will describe the decomposition of the four-point amplitude $A(1^+, 2^-, 3^+, 4^-)$. We know, of course, that the result should be given directly by the Parke-Taylor formula, since this is an MHV amplitude,

$$A(1^+, 2^-, 3^+, 4^-) = \frac{\langle 24\rangle^4}{\langle 12\rangle\langle 23\rangle\langle 34\rangle\langle 41\rangle} \tag{4.88}$$

We will verify that the decomposition procedure gives the same result. There are two BCFW graphs involved,



(4.89)

In order to calculate the corresponding expressions for the above graphs we need to find the expressions for the hatted momenta. In both the above graphs the same shifting vector is used

$$\epsilon_{a\dot{a}} = 4_a \tilde{1}_{\dot{a}} \tag{4.90}$$

where the notation

$$i_a \equiv \lambda_a(p_i) \qquad \tilde{i}_{\dot{a}} \equiv \tilde{\lambda}_{\dot{a}}(p_i) \tag{4.91}$$

is used. In what follows we might even suppress the dotted and undotted indices if the meaning is clear.

Moreover, the value of z used in both graphs is the same, since this value is found by the demand that the (common in both graphs) $\hat{P} = (p_1 \hat{+} p_2)$ quantity has vanishing square:

$$\hat{P}^2 = (p_1 + p_2 + z\epsilon)^2 = 2p_1 \cdot p_2 + 2z(p_1 + p_2)\cdot\epsilon = 0 \Rightarrow z = -\frac{\langle 12\rangle[12]}{\langle 24\rangle[21]} = -\frac{\langle 12\rangle}{\langle 42\rangle} \tag{4.92}$$

Hence, we have (from eq.(4.85))

$$\hat{p}_1 = 1\tilde{1} - \frac{\langle 12\rangle}{\langle 42\rangle}4\tilde{1} = 1\tilde{1} + \frac{\langle 24\rangle}{\langle 42\rangle}1\tilde{1} + \frac{\langle 41\rangle}{\langle 42\rangle}2\tilde{1} = \frac{\langle 41\rangle}{\langle 42\rangle}2\tilde{1} \tag{4.93}$$

where we have used the Schouten identity, eq.(4.43), to turn $\langle 12\rangle 4_a$ into $-\langle 24\rangle 1_a - \langle 41\rangle 2_a$.

Similarly

$$\hat{P} = q\tilde{q} + 2\tilde{2} - \frac{\langle 12\rangle}{\langle 42\rangle}4\tilde{1} = 2(\tilde{2} + \frac{\langle 41\rangle}{\langle 42\rangle}\tilde{1}) \tag{4.94}$$

and

$$\hat{p}_4 = 4\tilde{4} + \frac{\langle 12\rangle}{\langle 42\rangle}4\tilde{1} = 4(\tilde{4} + \frac{\langle 12\rangle}{\langle 42\rangle}\tilde{1}) \tag{4.95}$$

We are now ready to calculate the two graphs. For the first one we have

$$(a) = \frac{[\hat{1}\hat{P}]^3}{[\hat{P}2][2\hat{1}]}\frac{1}{(p_1+p_2)^2}\frac{\langle\hat{P}\hat{4}\rangle^3}{\langle\hat{4}3\rangle\langle 3\hat{P}\rangle} = \frac{[12]^3}{\frac{\langle 41\rangle}{\langle 42\rangle}[12][21]}\frac{1}{\langle 12\rangle[12]}\frac{\langle 24\rangle^3}{\langle 43\rangle\langle 32\rangle} = \frac{\langle 24\rangle^4}{\langle 12\rangle\langle 23\rangle\langle 34\rangle\langle 41\rangle} \tag{4.96}$$

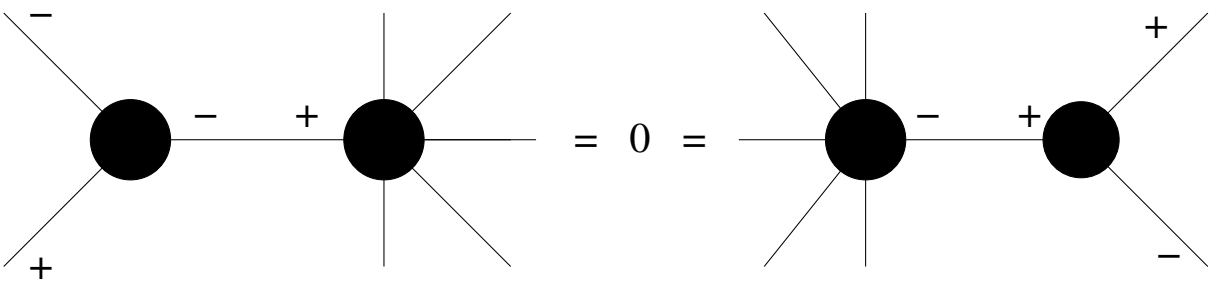
whereas the second graph vanishes

$$\frac{\langle 2\hat{P}\rangle^3}{\langle\hat{P}\hat{1}\rangle\langle\hat{1}2\rangle}\frac{1}{(p_1+p_2)^2}\frac{[\hat{P}3]^3}{[3\hat{4}][\hat{4}\hat{P}]} \tag{4.97}$$

due to $\langle 2\hat{P}\rangle^3 = \langle 22\rangle^3 = 0$.

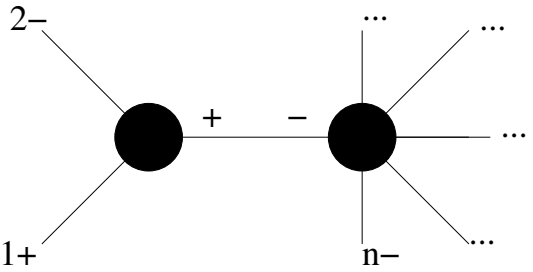
Reproducing MHV amplitudes

This phenomenon of vanishing three-leg subgraphs is widespread. In the framework of BCFW decompositions a three-leg subgraph will necessarily contain one of the two selected legs, the first or the last one, as well as the 'internal' leg carrying momentum $\hat{P}$. It is not difficult to see that if the third momentum has opposite helicity than the selected leg then the subgraph where the internal momentum has also opposite helicity than the selected leg vanishes.



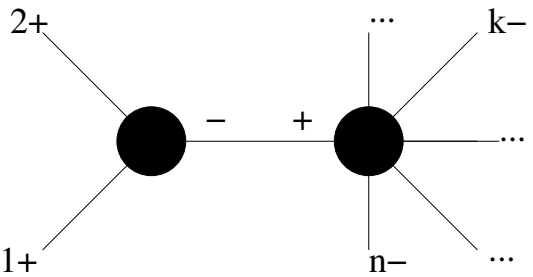
(4.98)

It is not hard to see that every MHV amplitude is trivially reproduced by the BCFW decomposition. A mostly plus amplitude, for example, has two external legs with negative helicity. In any BCFW decomposition where these two legs (of which one is the denoted as "last") are in different vertices, the corresponding graph vanishes (since the helicities of the internal propagator will force one of the two vertices to have only one negative helicity and hence to vanish). A possible exception to this would be the three-leg vertices which, being exceptional, can have only one negative helicity leg. This can actually be the contributing graph if the negative helicity leg is the one that follows the first leg in the cyclic ordering. Then one can have a graph of the type of fig.4.99. The other possibility, that the three-leg vertex involves the last leg (of negative helicity) is exactly the one, shown in fig.4.98, that vanishes.



(4.99)

Therefore, if the negative helicity legs are not the second and the last, they should be located at the same BCFW vertex. This forces the other vertex (the one including the leg denoted as "first") to have only three legs (otherwise it would vanish, having all but one leg with positive helicity). In that case, there is only one graph in the decomposition, the following:



(4.100)

Due to this, particular to MHV amplitudes, nicety it is not difficult to see that the BCFW decomposition reproduces any MHV amplitude in a transparent manner. For example the

seven-point amplitude, $A(1^+,2^+,3^+,4^-,5^+,6^+,7^-)$ can be written successively as

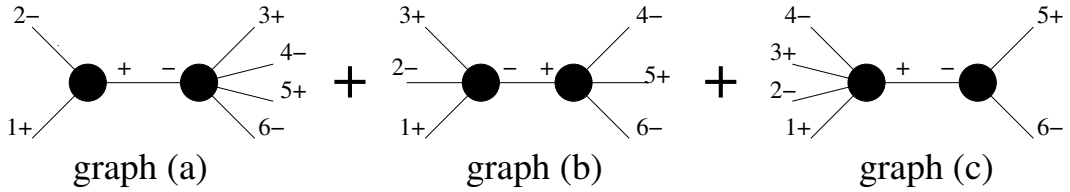
$$\begin{aligned}
A(1^+,2^+,3^+,4^-,5^+,6^+,7^-) &= \frac{\langle 72\rangle}{\langle 71\rangle\langle 12\rangle}A(2^+,3^+,4^-,5^+,6^+,7^-) = \\
&= \frac{\langle 72\rangle}{\langle 71\rangle\langle 12\rangle}\frac{\langle 73\rangle}{\langle 72\rangle\langle 23\rangle}A(3^+,4^-,5^+,6^+,7^-) = \\
&= \frac{\langle 72\rangle}{\langle 71\rangle\langle 12\rangle}\frac{\langle 73\rangle}{\langle 72\rangle\langle 23\rangle}\frac{\langle 74\rangle}{\langle 73\rangle\langle 34\rangle}A(4^-,5^+,6^+,7^-) = \\
&= \frac{\langle 74\rangle}{\langle 71\rangle\langle 12\rangle\langle 23\rangle\langle 34\rangle}A(5^+,6^+,7^-,4^-) = \\
&= \frac{\langle 74\rangle}{\langle 71\rangle\langle 12\rangle\langle 23\rangle\langle 34\rangle}\frac{\langle 46\rangle}{\langle 45\rangle\langle 56\rangle}A(6^+,7^-,4^-) = \\
&= \frac{\langle 74\rangle}{\langle 71\rangle\langle 12\rangle\langle 23\rangle\langle 34\rangle}\frac{\langle 46\rangle}{\langle 45\rangle\langle 56\rangle}\frac{\langle 74\rangle^3}{\langle 46\rangle\langle 67\rangle}
\end{aligned} \tag{4.101}$$

which is precisely the Parke-Taylor formula. In the above reduction the only input is the formula for the three-vertex $A(6^+,7^-,4^-)$.

Note that the corresponding reduction of a mostly minus MHV amplitude would be achieved by factoring out the last leg, n^-, first, thus yielding expressions with square brackets.

Beyond the MHVs

The promising mechanism of the previous section does not hold for amplitudes that are not MHVs, and the calculation has to be pursued in an orthodox way via eq.(4.84). The first class of non-MHV amplitudes involves six external legs. The $A(1^+,2^-,3^+,4^-,5^+,6^-)$, for example, can be decomposed into three graphs, as in fig.4.102.



(4.102)

To calculate graph C, as an example, we have to find the value of z determined by

$$\hat{P}^2 = (p_1+p_2+p_3+z\epsilon)^2 = 0 \rightarrow z = -\frac{P^2}{2P\cdot\epsilon} \tag{4.103}$$

Then we find that

$$\hat{P} = p_1+p_2+p_3-\frac{P^2}{2P\cdot\epsilon}\epsilon \tag{4.104}$$

$$\hat{p}_1 = p_1-\frac{P^2}{2P\cdot\epsilon}\epsilon = (1-\frac{P^2}{2P\cdot\epsilon}6)\tilde{1} \tag{4.105}$$

$$\hat{p}_6 = p_6+\frac{P^2}{2P\cdot\epsilon}\epsilon = 6(\tilde{6}+\frac{P^2}{2P\cdot\epsilon}\tilde{1}) \tag{4.106}$$

We can then calculate directly graph C as

$$A_c = \frac{\langle 2\hat{P}\rangle^4}{\langle\hat{1}2\rangle\langle 23\rangle\langle 3\hat{P}\rangle\langle\hat{P}\hat{1}\rangle}\frac{1}{P^2}\frac{\langle 4\hat{6}\rangle^4}{\langle\hat{6}\hat{P}\rangle\langle\hat{P}4\rangle\langle 45\rangle\langle 5\hat{6}\rangle} \tag{4.107}$$

Note that the intermediate, off-shell leg enters the expressions via its on-shell continuation, $\hat{P}$, as always, but cannot be written as a pair of two spinors in such a way that these two spinors

are simply related to the spinors of the external legs, as was the case in the previous section. Of course, it is always possible to express $\hat{P}$ as a pair of spinors. However, to do this one has to calculate the matrix for $\hat{P}$ and from that devise a form for the two spinors. This can only be done numerically. Instead, one can multiply every spinor product containing $\hat{P}$ with its complementary to arrive at an expression involving $\hat{P}$ itself instead of its spinors:

$$\begin{aligned}\langle i\hat{P}\rangle &= \frac{\langle i\hat{P}\rangle[1\hat{P}]}{[1\hat{P}]} = \frac{\lambda_a(p_i)\lambda_b(\hat{P})\epsilon^{ab}\tilde{\lambda}_{\dot{a}}(\hat{P})\tilde{\lambda}_{\dot{b}}(p_1)\epsilon^{\dot{b}\dot{a}}}{[1\hat{P}]} = \\ &= \frac{\lambda_a(p_i)\epsilon^{ab}\tilde{\lambda}_{\dot{b}}(p_1)\epsilon^{\dot{b}\dot{a}}\hat{P}_{b\dot{b}}}{[1\hat{P}]} \equiv \frac{\langle i|\hat{P}|1]}{[1\hat{P}]}\end{aligned} \tag{4.108}$$

Similarly

$$[\hat{P}j] = \frac{\langle n|\hat{P}|j]}{\langle n\hat{P}\rangle} \tag{4.109}$$

It is not hard to see that the extra factors of $\langle n\hat{P}\rangle$ and $[1\hat{P}]$ can only appear in the combination

$$\langle n\hat{P}\rangle[1\hat{P}] = \langle n|\hat{P}\ 1] = \sum_i \langle ni\rangle[i1] \quad P = p_1 + \ldots + p_i \tag{4.110}$$

in any BCFW graph.

The exact expressions for the six-point amplitude are not of importance here. They can be found in [7] where it is shown that they agree with the previously obtained expressions for this particular color-amplitude. The final expressions for non-MHV color-ordered amplitudes are by no means simple, but they are considerably simpler than intermediate expressions in the calculation thanks to various cancelations that are taking place. The BCFW decomposition reproduces the simpler final results in a very non-trivial way, relying solely on spinor identities (like the Schouten identity) and momentum conservation. The way the correct result is obtained is by no means transparent, and would actually appear to be rather miraculous to anybody familiar with such kind of algebraic manipulations. Yet, results obtained by the BCFW decomposition have been numerically cross-checked for amplitudes with up to 12 external legs!

It should be clear by now that the BCFW formalism has considerable advantages as a tool for analytic calculations, over the previously used techniques (like the spinor techniques [41] or the Berends-Giele recursion relations [7]). It is equally clear that this formalism has no straightforward interpretation in terms of Feynman diagrams. For what matters, it seems that any four-vertex is eliminated from the theory, at least in the diagrammatic sense, and that all amplitudes are reduced down to the elementary and exceptional three-point vertices.

4.5.3 The BCFW proof

Britto, Cachazo, Feng and Witten gave a proof [14] of the BCFW recursion relations that relied heavily on analyticity properties of the color amplitudes. Although it turned the attention of the community back to the pole structure of such amplitudes[11], the proof shed little light on the connection between the BCFW formalism and the familiar Feynman diagram approach. We briefly present this proof below.

Let's consider a general color-ordered amplitude $A(p_1, p_2, \ldots, p_{n-1}, p_n)$ and define a complex function of z

$$A(z) = (\hat{p}_1, p_2, \ldots, p_{n-1}\hat{p}_n) \tag{4.111}$$

Three facts are claimed for this function.

[11] This is a twist that has been in fashion in the '60s, under the name of the 'bootstrap' program, whose goal was to reconstruct the whole scattering amplitude directly from its analyticity properties, see [23].

At first, this is obviously a rational function in z since z enters the expression through $\tilde{\lambda}_{\dot{a}}(p_n)$ and $\lambda_a(p_1)$ linearly.

Secondly, possible poles in z can only be single poles: since the amplitude is the sum of ordered Feynman diagrams, z enters the denominator of each diagram through its propagators. Since we are only considering tree-level amplitudes, each propagator is raised to the unit power - there are no propagators raised to a higher power. Note incidentally that for a propagator to contribute a pole in z it needs to carry one but not both of the special momenta $\hat{p}_1, \hat{p}_n$, i.e., it needs to lie along the line of propagators that connect the two special legs.

Finally, it is claimed that the limit of the function $A(z)$ as $z \to \infty$ goes to zero.

$$\lim_{z\to\infty} A(z) = 0 \tag{4.112}$$

This has been justified in [14] as follows: the value of the limit depends on the balance of powers of z in the numerator and the denominator of $A(z)$. The numerator can only get contributions from three-gluon vertices which carry momentum. Four-gluon vertices do not contribute any powers of z. Moreover, a three-gluon vertex contributes a single power of z if and only if it lies along the line of propagators that connect the first and the last leg. Assume that there are R propagators between the first and the last leg for the particular Feynman diagram in consideration. The maximum possible number of three-gluon vertices is $R+1$, hence the numerator goes at most as z^{R+1}. Every propagator along that line also contributes a power of z in the denominator of $A(z)$. Hence, overall, the vertices and propagators give a factor that grows linearly with z in the worse behaving Feynman diagram. The claim, then, is that the two polarization vectors for the two selected legs contribute a factor of $\frac{1}{z}$ each, thus setting the overall behavior of $A(z)$ as $\frac{1}{z}$, i.e. ensuring that as $z \to \infty$ the function $A(z)$ goes to zero.

This last claim is actually a gauge-dependent statement. One can define a particular gauge for the polarization vectors of the selected external gluons in which they are actually independent of z. Indeed, we shall do so in the following sections. In that gauge the limit of $A(z)$ as $z \to \infty$ does go to zero, but the limit of individual graphs does not. One can always appeal to gauge invariance in order to restore the proof that $A(z)$ goes to zero at infinity. This will turn out to be an instrumental observation for the diagrammatic proof of the BCFW decomposition that we will give later.

Given the validity of the above claims, we can write $A(z)$ as a sum over its residues

$$A(z) = \sum_i \frac{c_i}{z - z_i} \tag{4.113}$$

where z_i are the poles of $A(z)$. These poles can only appear due to the propagators along the line from the first to the last leg [12], so we can actually write $z_i = -\frac{P_i^2}{2P_i\epsilon}$

$$A(z) = \sum_{i=1}^{R} \frac{c_i}{z + \frac{P_i^2}{2P_i\epsilon}} = \sum_{i=1}^{R} \frac{c_i(2P_i\cdot\epsilon)}{2P_i\cdot\epsilon z + P_i^2} \equiv \sum_{i=1}^{R} \frac{r_i}{\hat{P}_i^{\,2}} \tag{4.114}$$

where P_i is the momentum carried by this propagator and $r_i \equiv c_i 2P_i\cdot\epsilon$.

The residues, r_i are found as follows: a pole z_i from the propagator carrying momentum $P_i = p_1 + p_2 + \ldots + p_i$ appears due to all possible Feynman diagrams that actually have such a propagator along the line that connects the first and the last legs. Each such graph is naturally split in a 'left' part containing the first i legs and a 'right' part containing the remaining $n-i$ legs. The sum of these parts is seen to be the full $i+1$- and $n-i+1$-point amplitude with one off-shell leg respectively and with $\hat{p_1}$ and $\hat{p_n}$ instead of p_1 and p_n. The propagator itself can have its positive helicity end attached to the 'left' or the 'right' part. To evaluate the residue

[12] This is true only when the polarization vectors of the external legs are actually z-independent!

we have to set $z = z_i$ in the 'left' and 'right' parts. When this is done, the two parts become on-shell amplitudes with two complex momenta.

Therefore, the residue at $z = z_i$ is

$$r_i = \sum_{h=\pm} \left[A(\hat{p}_1, p_2, \ldots, p_i, \hat{P}_i^{\,h}) \right]_{z=z_i} \frac{1}{\hat{P}_i^{\,2}} \left[A(-\hat{P}_i^{\,-h}, p_{i+1}, \ldots, p_{n-1}, \hat{p}_n) \right]_{z=z_i} \tag{4.115}$$

which gives

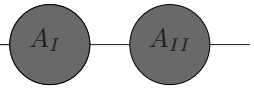
$$A(p_1, \ldots, p_n; z) = \sum_{h=\pm} \left[A(\hat{p}_1, p_2, \ldots, p_i, \hat{P}_i^{\,h}) \right]_{z=z_i} \frac{1}{\hat{P}_i^{\,2}} \frac{1}{\hat{P}_i^{\,2}} \left[A(-\hat{P}_i^{\,-h}, p_{i+1}, \ldots, p_{n-1}, \hat{p}_n) \right]_{z=z_i} \tag{4.116}$$

The BCFW decomposition formula eq.(4.84) is equivalent to the above one for $z = 0$.

4.5.4 Counting contributions

The BCFW decomposition is a recursive decomposition: every n-point dual amplitude is decomposed into smaller amplitudes, which in turn can be further decomposed into even smaller amplitudes and so on, until the special three-point amplitude which is not decomposable. The result of such decompositions, in terms of graphs, are fully decomposed diagrams that contain exclusively three-point vertices[13]. The question immediately arises: is there a relation between these fully decomposed graphs and the usual Feynman graphs? To begin with, are these two types of diagrams equal in number? To find out, one has to count BCFW diagrams.

This is easily achieved if we neglect any reference to polarization vectors or momenta in the BCFW formula, while keeping the ordering of the external legs. In that case of course we count three-vertex planar graphs (see [64]), and there is nothing to distinguish between BCFW fully decomposed graphs and Feynman graphs. In fact the equation itself can be used to count the graphs in a recursive manner. The form of the equation is given by

$\mathcal{A} = \sum$ 

(4.117)

or in a more mathematical form

$$A^{BCF}_{1\to n} = \sum_{k=1}^{n-3} A^{BCF}_{1\to k+1} A^{BCF}_{1\to n-k} \tag{4.118}$$

where $A^{BCF}_{1\to n}$ stands here for the number of graphs with $n+1$ legs. To give an example of how this works, let's examine the case $n = 4$. We have

$$A^{BCF}_{1\to 4} = A_{1\to 2} A^{BCF}_{1\to 3} + A^{BCF}_{1\to 3} A^{BCF}_{1\to 2} \tag{4.119}$$

and $A^{BCF}_{1\to 3}$ can be further decomposed to

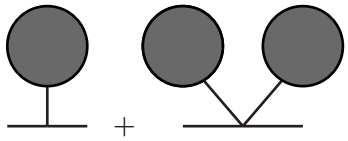
$$A^{BCF}_{1\to 3} = A^{BCF}_{1\to 2} A^{BCF}_{1\to 2} \tag{4.120}$$

where $A^{BCF}_{1\to 2} = 1$ since there is only one diagram with three legs. This gives $A^{BCF}_{1\to 3} = 1$ and $A^{BCF}_{1\to 4} = 2$.

It soon becomes obvious that the number of fully decomposed BCFW diagrams cannot be equal to the number of Feynman diagrams.

[13] Which should not to be confused with the familiar QCD three-vertices.

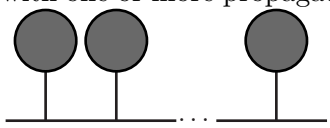
Firstly, diagrams where no propagator line exists between the 1st and the nth particle, are not included in the BCFW equation

$A^0 =$  (4.121)

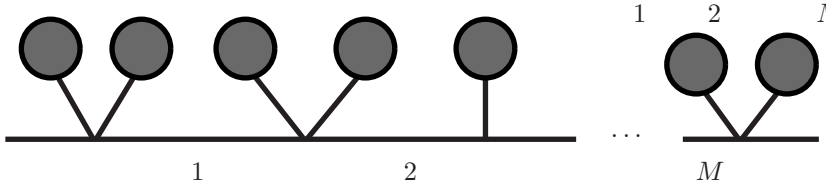
These diagrams are given by a recursive relation of the form

$$A^0_{1\to n} = A_{1\to n-1} + \sum_{n_1+n_2=n-1} A_{1\to n_1} A_{1\to n_2} \tag{4.122}$$

Secondly, BCFW is multiple-counting diagrams with one or more propagators along the line that connects the first with the last leg, such as

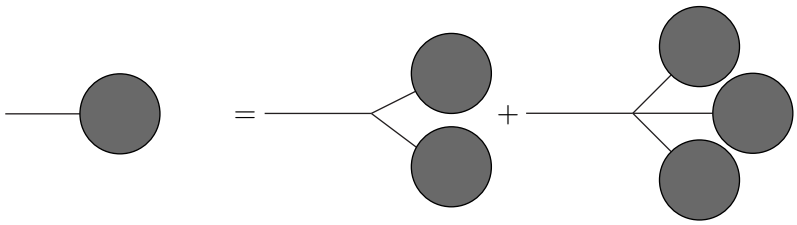


or



In fact, each diagram is counted exactly as many times as the number of propagators along the line.

To make our arguments more quantitative, let us begin with the Berends-Giele [5] (or Dyson-Schwinger for ordered graphs [2], [27]) recursive equation for a generic theory with three- and four-vertices, like QCD.

 (4.123)

$$A_{1\to n} = \sum_{n_1+n_2=n} A_{1\to n_1} A_{1\to n_2} + \sum_{n_1+n_2+n_3=n} A_{1\to n_1} A_{1\to n_2} A_{1\to n_3} \tag{4.124}$$

Now, $A_{1\to n}$ counts Feynman diagrams with $n+1$ external legs. Then, the claim is that the number of Feynman diagrams relates to the number of fully decomposed BCFW diagrams by

$$\mathcal{A}_{1\to n} = \mathcal{A}^{\mathcal{BCF}}{}_{1\to n} + \mathcal{A}^0_{1\to n} - D \tag{4.125}$$

where $A^0_{1\to n}$ counts the diagrams in eq.(4.121) and the subtracted term D accounts for the overcounting of diagrams. Since the overcounting relates to diagrams with $2, 3, \ldots$ propagators on the line between the first and the last leg, it can be seen that

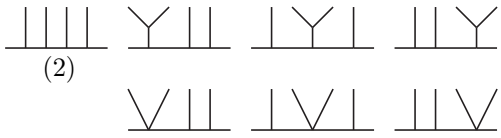
$$D = \sum_{M=3}^{n-1} (M-2) \sum_{k=0}^{M} \begin{pmatrix} M \\ k \end{pmatrix} D^n_{M+k} \tag{4.126}$$

where $M-1$ is the number of propagators of the particular overcounted class of diagrams and

$$D^n_M = \sum_{n_1+\ldots+n_M=n} A_{1\to n_1} \ldots A_{1\to n_M} \tag{4.127}$$

counts the number of diagrams within that class.

As an example for $n = 5$ (6-leg gluon amplitude), the number of Feynman graphs is $A = 38$, the number of BCFW graphs is $A^{BCF} = 29$, the number of graphs without any propagator between the first and the last leg is $A^0 = 17$ so overcounting should give $D = 8$. The following graphs are overcounted



(4.128)

with the first diagram doubly overcounted (it contains 3 propagators). That diagram corresponds to the $M = 4$ while the other graphs come from the $M = 3$ term of eq.(4.126).

In the table below we give the results for up to $n = 11$ particles.

$1 \to n$	A	A^{BCF}	A^0	$A^{BCF} + A^0 - A$	D
3	3	1	2	0	0
4	10	6	5	1	1
5	38	29	17	8	8
6	154	136	64	46	46
7	654	636	259	241	241
8	2871	2992	1098	1219	1219
9	12925	14190	4815	6080	6080
10	59345	67860	21659	30174	30174
11	276835	327080	99385	149630	149630

This analysis strongly suggests that a connection between Feynman diagrams and the BCFW decomposition might be achieved by grouping together BCFW (hatted) diagrams with the same chain structure along the main line (the same number of propagators hence the same multiplicity) but differently placed cuts: the multiplicity of each group is equal to the number of propagators along the main line which in turns equals the number of possible cuts (one for each propagator along the main line).

4.5.5 Hatted functions and some kinematical identities

In this section we will slightly digress to define the 'hat' operation and derive some important but fairly general kinematical identities that will be abundantly used in what follows.

The hat symbol over a function of four-momenta $\hat{f}(q)$ will denote the function $\hat{f}(q; z)$ where the argument of f is analytically continued (shifted) by a four-vector $z\epsilon_\mu$, with

$$\epsilon_\mu \equiv \frac{1}{2}\overline{u}_-(p_1)\gamma_\mu u_-(p_n) = \frac{1}{2}\tilde{\lambda}_{\dot{a}}(p_1)\overline{\sigma}_\mu^{\dot{a}a}\lambda_a(p_n) \tag{4.129}$$

This has the effect

$$q \to \hat{q} = q + z\epsilon \tag{4.130}$$

We will omit the explicit reference to the z-dependence of $\hat{f}(p)$ since this is signified by the hat symbol.

For any function $\hat{f}(p; z)$ of z that has only simple poles in z and vanishes at $z \to \infty$, we can perform an expansion over its poles. In particular we have

$$\hat{f}(p; z) = \sum_j \Big[\hat{f}(p; z)(z - z_j)\Big]_{z=z_j} \frac{1}{z - z_j} \tag{4.131}$$

where the sum is over all the simple poles z_j of $\hat{f}(p;z)$. This very general identity allows us to analytically continue the function $f(p)$ to the complex plane, make use of the pole expansion and take the limit $z \to 0$ to return to the real axis, thus obtaining a relation between $f(p)$ and the pole expansion of $\hat{f}(p;z)$.

Applying this to a momentum antenna we get

$$\frac{1}{\hat{p}_1^2\hat{p}_2^2\dots\hat{p}_k^2} = \sum_{j=1..k}\left[\frac{1}{\hat{p}_1^2\hat{p}_2^2\dots\hat{p}_{j-1}^2\hat{p}_{j+1}\dots\hat{p}_k^2}\right]_{z=z_j}\frac{1}{\hat{p}_j^2} \tag{4.132}$$

with z_j such that

$$\hat{p}_j(z_j)^2 = (p_j + z_j\epsilon)^2 = p_j^2 + 2z_j p_j\cdot\epsilon = 0 \tag{4.133}$$

where $p_1\dots p_k$ are arbitrary off-shell four-momenta. In the next sections we will use the above identity with p_i being sums of on-shell momenta of the form $p_{1k} = p_1 + p_2 + \dots + p_k$.

Taking the limit $z \to 0$ at both sides we have

$$\frac{1}{p_1^2p_2^2\dots p_k^2} = \sum_{j=1..k}\left[\frac{1}{\hat{p}_1^2\hat{p}_2^2\dots\hat{p}_{j-1}^2\hat{p}_{j+1}^2\dots\hat{p}_k^2}\right]_{z=z_j}\frac{1}{p_j^2} \tag{4.134}$$

Further more,

$$\frac{z^\rho}{\hat{p}_1^2\hat{p}_2^2\dots\hat{p}_k^2} = \sum_{j=1..k}\left[\frac{z^\rho}{\hat{p}_1^2\hat{p}_2^2\dots\hat{p}_{j-1}^2\hat{p}_{j+1}^2\dots\hat{p}_k^2}\right]_{z=z_j}\frac{1}{\hat{p}_j^2} \tag{4.135}$$

which gives the very useful set of identities, valid for every $\rho < k$

$$\sum_{j=1..k}\left[\frac{z^\rho}{\hat{p}_1^2\hat{p}_2^2\dots\hat{p}_{j-1}^2\hat{p}_{j+1}\dots\hat{p}_k^2}\right]_{z=z_j}\frac{1}{p_j^2} = 0 \tag{4.136}$$

Finally, if $\rho = k$, the function in the left hand side of eq.(4.135) is no longer vanishing at $z \to \infty$. Subtracting its limit at infinity, however, we have a new function that does, so

$$\frac{z^k}{\hat{p}_1^2\hat{p}_2^2\dots\hat{p}_k^2} - \lim_{z\to\infty}\frac{z^k}{\hat{p}_1^2\hat{p}_2^2\dots\hat{p}_k^2} = \sum_{j=1..k}\left[\frac{z^k}{\hat{p}_1^2\hat{p}_2^2\dots\hat{p}_{j-1}^2\hat{p}_{j+1}\dots\hat{p}_k^2}\right]_{z=z_j}\frac{1}{\hat{p}_j^2} \tag{4.137}$$

and taking the limit $z \to 0$ we get

$$-\lim_{z\to\infty}\frac{z^k}{\hat{p}_1^2\hat{p}_2^2\dots\hat{p}_k^2} = \sum_{j=1..k}\left[\frac{z^k}{\hat{p}_1^2\hat{p}_2^2\dots\hat{p}_{j-1}^2\hat{p}_{j+1}\dots\hat{p}_k^2}\right]_{z=z_j}\frac{1}{p_j^2} \tag{4.138}$$

or

$$\frac{-1}{\prod_{j=1}^k 2\epsilon\cdot p_j} = \sum_{j=1..k}\left[\frac{z^k}{\hat{p}_1^2\hat{p}_2^2\dots\hat{p}_{j-1}^2\hat{p}_{j+1}\dots\hat{p}_k^2}\right]_{z=z_j}\frac{1}{p_j^2} \tag{4.139}$$

4.5.6 Choosing a gauge

We have seen in section 4.5.4 that classes of Feynman diagrams should somehow correspond to particular BCFW decompositions. Using a particular gauge for the external gluons one can eliminate whole classes of Feynman diagrams.

A consistent definition of the external gluons' polarization vectors was given in eq.(4.59) where $\overline{p}$ is an auxiliary, null four-vector that can be chosen at will for every gluon as long as it is not parallel to p itself.

We choose to use $\overline{p}_1 = p_n$ and $\overline{p}_n = p_1$

$$\epsilon^{\mu}_{+1} = \frac{\overline{u}_{+n}\gamma^{\mu}u_{+1}}{\sqrt{2}\overline{u}_{+n}u_{-1}} = \frac{\tilde{\lambda}_{\dot{a}}(p_1)\overline{\sigma}^{\mu,\dot{a}a}\lambda_a(p_n)}{\sqrt{2}\langle n1\rangle} \tag{4.140}$$

$$\epsilon^{\mu}_{-n} = \frac{\overline{u}_{-1}\gamma^{\mu}u_{-n}}{\sqrt{2}\overline{u}_{-n}u_{+1}} = -\frac{\tilde{\lambda}_{\dot{a}}(p_1)\overline{\sigma}^{\mu,\dot{a}a}\lambda_a(p_n)}{\sqrt{2}\,[n1]} \tag{4.141}$$

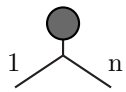
A number of simplifications follow immediately. The product of ϵ_1 with ϵ_n or ϵ (defined in eq.(4.129)) vanishes

$$\boxed{\epsilon_1\cdot\epsilon_n = 0 = \epsilon\cdot\epsilon_1 = \epsilon\cdot\epsilon_n} \tag{4.142}$$

Moreover

$$\epsilon_{+1}\cdot p_n = 0 = \epsilon_{-n}\cdot p_1. \tag{4.143}$$

As a consequence, any diagram in which the first and the last leg meet in a three-vertex vanishes.



$$= 0 \tag{4.144}$$

where the blob denotes any Feynman diagram with the particular off-shell leg.

These simplifications are in the correct direction in view of the fact that there are no BCFW graphs with the first and the last leg in the same vertex. We still have to deal with the case when 1 and n are attached in a four-vertex whose two other lines are off-shell (another situation that doesn't occur in a BCFW decomposition). We will see in the next section how to accommodate these diagrams.

Finally, we should note that the polarization vectors ϵ^{μ}_{+1} and ϵ^{μ}_{-n} actually differ by a complex phase. They also differ by a complex factor from the shifting vector ϵ (see 4.129). Note that in any expression where all these three vectors appear in scalar products, one is allowed to freely interchange them, without altering the result. In a very real sense there is actually only one polarization vector in this gauge.

Finally, the polarization vectors are now invariant under the shifting operation:

$$p_1 \to p_1 + z\epsilon \quad p_n \to p_n - z\epsilon \tag{4.145}$$

If we write

$$p^{\mu}_1 = \frac{1}{2}\overline{u}_{-1}\gamma^{\mu}u_{-1} \to \frac{1}{2}\overline{u}_{-1}\gamma^{\mu}u_{-1} + \frac{1}{2}z\overline{u}_{-1}\gamma^{\mu}u_{-n} \tag{4.146}$$

and

$$p^{\mu}_n = \frac{1}{2}\overline{u}_{-n}\gamma^{\mu}u_{-n} \to \frac{1}{2}\overline{u}_{-n}\gamma^{\mu}u_{-n} + \frac{1}{2}z\overline{u}_{-1}\gamma^{\mu}u_{-n} \tag{4.147}$$

we see that the shifting operator effectively sends

$$\begin{aligned} u_{-1} &\to u_{-1} + zu_{-n} \\ \overline{u}_{-n} &\to \overline{u}_{-n} - z\overline{u}_{-1} \end{aligned} \tag{4.148}$$

or, in terms of Weyl - Van der Waerden spinors

$$\begin{aligned} \lambda_a(p_1) &\to \lambda_a(p_1) + z\lambda_a(p_n) \\ \tilde{\lambda}_{\dot{a}}(p_n) &\to \tilde{\lambda}_{\dot{a}}(p_n) - z\tilde{\lambda}_{\dot{a}}(p_1) \end{aligned} \tag{4.149}$$

As a result, the denominators of ϵ_{+1} and ϵ_{-n} become

$$\langle n1\rangle \to \langle n1\rangle + z\langle nn\rangle = \langle n1\rangle \tag{4.150}$$

and

$$[n1] \to [n1] - z[11] = [n1] \tag{4.151}$$

The only restriction that we impose on the polarization vectors of the other gluons is that they remain invariant under the shifting operator. In case their assisting vectors $\overline{p}_i$ involve p_1 or p_n this should be arranged in a way that preserves this invariance.

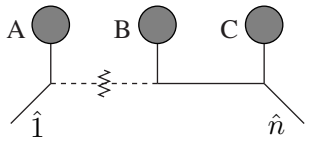
This gauge choice explicitly eliminates any z-dependence from the polarization vectors. The shifting operation affects amplitudes only through the shift at the momentum four-vectors p_1 and p_n, and the induced shift to the momenta of intermediate propagators. Polarization vectors and vertices that do not carry p_1 and p_n are left unchanged. Therefore, a diagram after shifting will be a complex function with poles coming from three-vertices or propagators.

This is in contrast with a general gauge, where the polarization vectors ϵ_{+1} and ϵ_{-n} get an extra factor involving z in their denominator. A pole expansion, in that case, would need to take into account (gauge-dependent) poles coming from the gluon polarization vectors.

Counting powers of z in the general gauge shows (see the concluding discussion in [14]) that any diagram vanishes at the limit $z \to \infty$. This is *n*ot, in general, true in the gauge we are working. The $z \to \infty$ limit of a diagram is used as a guide to group diagrams in classes where these limits cancel.

4.5.7 Hatted graphs

Let us call 'hatted' diagrams graphs of the form



(4.152)

denoting the corresponding Feynman graph where we have multiplied by p_{1A}^2, performed the momenta shift, evaluated it at some $z = z_i$ and divided back by p_{1A}^2. The value z_i is defined by the demand that the 'cut' propagator, carrying momentum $\hat{p}_{1A}$, vanishes: $\hat{p}_{1A}^2 = 0$.

In terms of 'hatted' graphs, the BCFW decomposition consists in two blobs, one containing the first and the other the last leg, the amplitude is hatted and a cut is taken on the propagator that connects the two blobs. A sum over all partitions of n in two integers is employed. Evidently, every particular BCFW decomposition over some propagator P_i is equivalent to the sum of all hatted diagrams with that particular propagator cut.

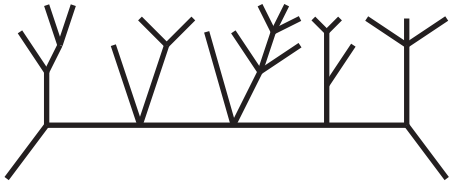
Thus the sum of all possible decompositions is equivalent to the sum of all hatted diagrams with all possible cuts along the line of propagators that connects the first and the last leg of the original color amplitude.

In what follows we will prove that the sum of all possible hatted diagrams is equivalent to the sum of Feynman diagrams involved in the computation of the particular amplitude.

4.5.8 The correspondence of hatted graphs to Feynman diagrams

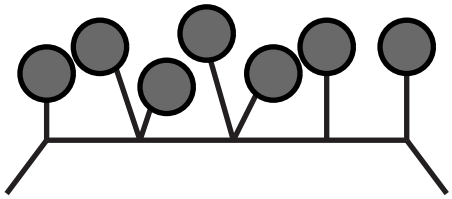
The special treatment of the two selected external legs 1 and n implies a special classification of the various Feynman diagrams involved in the given amplitude. There are classes of diagrams with zero, one,two,... propagators on the line between the first and the last leg. Every class consists of subclasses defined by the particular partition of the other $n-2$ external legs in which they are connected to the main line.

A particular diagram with $n = 20$ would be, for example,



(4.153)

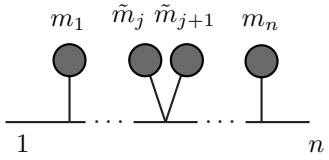
and would correspond to the class



(4.154)

and more precisely to its subclass with $3, 1, 2, 1, 5, 3, 3$ on-shell legs in the corresponding blobs[14]. Let $m_1, m_2, \ldots, m_{k+1}$ be a particular partition of $n-2$, such that $\sum m_i = n-2$. The blobs denote the full set of all possible color-ordered diagrams with one off-shell and m_i on-shell legs.

Let us call this subclass of Feynman diagrams $F(m_1, \ldots, \tilde{m}_j, \tilde{m}_{j+1}, \ldots, m_{k+1})$ where the tildes on top of two consecutive m's mark the presence of a four-vertex attaching the corresponding blobs to the line.

$F(m_1, \ldots, \tilde{m}_j, \tilde{m}_{j+1}, \ldots, m_{k+1}) =$ 

(4.155)

The particular subclass of diagrams in eq.(4.154) would then be called $F(3, \tilde{1}, \tilde{2}, \tilde{1}, \tilde{5}, 3, 3)$ and eq.(4.153) would be one of the actual Feynman diagrams contained in that subclass.

In the gauge we are working, the hatting operator would leave the contribution of the blobs, in such a diagram, invariant, since the momentum flowing into the blob doesn't contain p_1 or p_n. It would only affect the propagators and vertices along the main line. After the shifting of the momenta a cut can be placed on any of the propagators on the line, thus fixing the value of z at which the residue should be evaluated. We shall denote such a graph with a cut on the j'th propagator by

$$H(m_1, \ldots, m_{k+1}; j) = \left[\hat{F}(m_1, \ldots, m_{k+1})\hat{p}_{1j}^2\right]_{z=z_i} \frac{1}{p_{1j}^2} \tag{4.156}$$

Here and in what follows p_{1A_i} denotes the sum of external momenta

$$p_{1A_i} = p_1 + p_{A_i} + p_{A_2} + \ldots + p_{A_i} \tag{4.157}$$

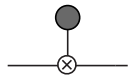
with

$$p_{A_i} = \sum_{k=1}^{m_i} p_k \tag{4.158}$$

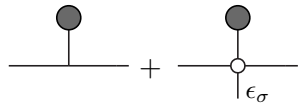
A closer look at the "hatted" diagram will reveal that hatting affects, through the momenta p_1 and p_n, only the three-vertices along the main line. The four-vertices, not carrying any momentum, remain unchanged. The effect on the three-vertices is included in the "hatted" YM

[14] A blob with m_i on-shell legs and one off-shell denotes the corresponding sum of color-ordered diagrams with one off-shell leg and m_i on-shell legs.

three-gluon vertices, which we denote with a crossed white blob in the figures. Each of these can be decomposed in two pieces:

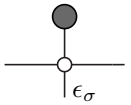


$$
\begin{aligned}
&= \hat{V}_{\mu\nu\rho} = g_{\mu\nu}(\hat{p}_1 - p_2)_\rho + g_{\nu\rho}(p_2 - \hat{p}_3)_\mu + g_{\rho\mu}(\hat{p}_3 - \hat{p}_1)_\nu \\
&= g_{\mu\nu}(p_1 - p_2)_\rho + g_{\nu\rho}(p_2 - p_3)_\mu + g_{\rho\mu}(p_3 - p_1)_\nu \\
&\quad -z\epsilon^\sigma(2g_{\mu\rho}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\rho} - g_{\mu\nu}g_{\sigma\rho}) \\
&= V_{\mu\nu\rho} - z\epsilon^\sigma(2g_{\mu\rho}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\rho} - g_{\mu\nu}g_{\sigma\rho})
\end{aligned}
$$



(4.159)

where



$$= -zV_{\mu\nu\rho\sigma} \tag{4.160}$$

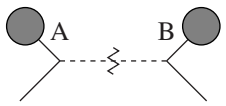
with $V_{\mu\nu\rho\sigma}$ the QCD four-vertex.

Let us concentrate for the moment on the class of diagrams with the smallest number of propagators along the line between the first and the last leg: diagrams with no propagators along the line can have a three or a four-vertex joining these two legs. The whole class of diagrams with a three-vertex joining the first and the last leg vanishes identically in the gauge we are using. The remaining class with one four-vertex will be dealt with soon.

Next, we have the class with one propagator along the main line.

If we write J^μ_Q for the current coming from the blob Q, we have

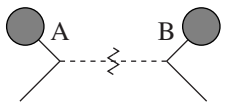
$$
\begin{aligned}
H(m_1, m_2; 1) \;\equiv\; & \\
&= \Big[\epsilon_{1\mu} J_{A\nu} \hat{V}^{\mu\nu\rho} \hat{V}_{\rho\kappa\lambda} J^\kappa_B \epsilon^\lambda_n\Big]_{z_0} \frac{1}{p^2_{1A}} \\
&= \Big[\epsilon_{1\mu} J_{A\nu} (V^{\mu\nu\rho} - z\epsilon_\sigma V^{\sigma\mu\nu\rho})(V_{\rho\kappa\lambda} - z\epsilon^\tau V_{\tau\rho\kappa\lambda}) J^\kappa_B \epsilon^\lambda_n\Big]_{z_0} \frac{1}{p^2_{1A}} \\
&= \Big[\epsilon_{1\mu} J_{A\nu} V^{\mu\nu\rho} V_{\rho\kappa\lambda} J^\kappa_B \epsilon^\lambda_n - z\epsilon_{1\mu} J_{1\nu} \epsilon_\sigma V^{\sigma\mu\nu\rho} V_{\rho\kappa\lambda} J^\kappa_B \epsilon^\lambda_n \\
&\quad -z\epsilon_{1\mu} J_{A\nu} V^{\mu\nu\rho} \epsilon^\tau V_{\tau\rho\kappa\lambda} J^\kappa_B \epsilon^\lambda_n \\
&\quad +z^2 \epsilon_{1\mu} J_{A\nu} \epsilon_\sigma V^{\sigma\mu\nu\rho} \epsilon^\tau V_{\tau\rho\kappa\lambda} J^\kappa_B \epsilon^\lambda_n\Big]_{z_0} \frac{1}{p^2_{1A}}
\end{aligned}
$$



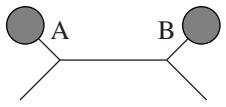
(4.161)

The first, z-independent term in the bracket equals the Feynman diagram with one propagator. The last term in the bracket vanishes due to eq.(4.142) and we get

$$H(m_1, m_2; 1) \;\equiv$$



$$=$$

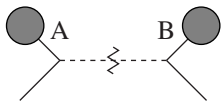


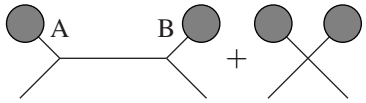
$$- [2(\epsilon\cdot p_{1A})(\epsilon_1\cdot J_A)(\epsilon_n\cdot J_B)]_{z_0} \frac{1}{p^2_{1A}}$$

with

$$z_0 = -\frac{p_{1A}^2}{2(p_{1A}\epsilon)} \tag{4.162}$$

where we have used momentum conservation and our freedom to interchange the polarization vectors. So

$H(m_1, m_2; 1) \equiv$ 

$=$ 
 (4.163)

or

$$H(m_1, m_2; 1) = F(m_1, m_2) + F(\tilde{m}_1, \tilde{m}_2) \tag{4.164}$$

In the case that the blobs A and B of the first diagram in eq.(4.163) contain exactly one on-shell leg, the BCFW decomposition leads to two hatted QCD 3-vertices. We have verified that the usual QCD 3-vertex evaluated at the hatted kinematics reproduces the formulas

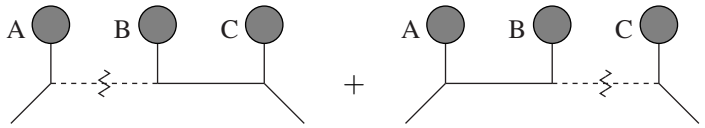
$$A(\hat{1}^+, 2^-, \hat{q}^+) = \sqrt{2}\frac{[\hat{q}\hat{1}]^3}{[\hat{1}2][2\hat{q}]} \tag{4.165}$$

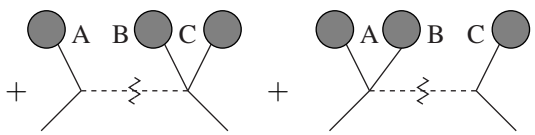
and

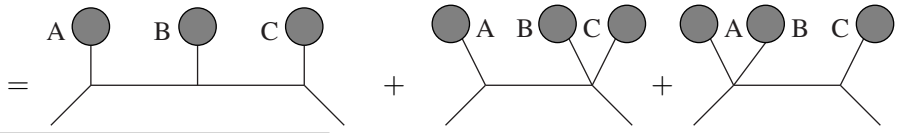
$$A(\hat{1}^+, 2^+, \hat{q}^-) = \sqrt{2}\frac{[\hat{1}\hat{2}]^3}{[2\hat{q}][\hat{q}1]} \tag{4.166}$$

where $p_q = -p_1 - p_2$ is the off-shell leg and $\hat{p}_q = -p_1 - p_2 - z\epsilon$ is its on-shell continuation [15]. The corresponding formulas involving the last leg (which has negative helicity) work similarly but include angle bracket spinor products.

Proceeding to graphs that contain two propagators between the first and the last leg or graphs that contain one propagator with one four-vertex attached to it, we have

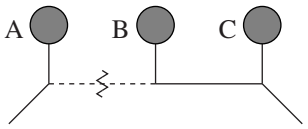

 (4.167)




 (4.168)

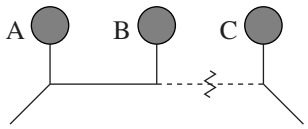
[15] The other polarizations $(A(\hat{1}^+, 2^+, \hat{q}^+), A(\hat{1}^+, 2^-, \hat{q}^-))$ can be shown to vanish.

The expressions that correspond to the first two hatted graphs on the left hand side are quadratic in z and are evaluated in different values of z, z_1 and z_2.



$$= \left[H_0 \frac{1}{\hat{p}^2_{1B}} + H_1 \frac{z}{\hat{p}^2_{1B}} + H_2 \frac{z^2}{\hat{p}^2_{1B}} \right]_{z=z_1} \frac{1}{p^2_{1A}} \tag{4.169}$$

with $p_{1B} = p_1 + p_A + p_B$



$$= \left[H_0 \frac{1}{\hat{p}^2_{1A}} + H_1 \frac{z}{\hat{p}^2_{1A}} + H_2 \frac{z^2}{\hat{p}^2_{1A}} \right]_{z=z_2} \frac{1}{p^2_{1B}} \tag{4.170}$$

The z-independent term of all four graphs gives the Feynman graphs of the right hand side of eq.(4.168), due to the kinematic identity eq.(4.134).

Moreover, the part linear in z of the first two graphs cancels identically due to eq.(4.136) with $\rho = 1$. The quadratic part of the first two graphs exactly cancels the part linear in z of the two last graphs on the left hand side of eq.(4.168).

The generalization of the above mechanism goes as follows: each hatted diagram with v_3 three-vertices along the main line can be decomposed in a sum of sub-diagrams having 0,1, ..., v_3 white-blob vertices corresponding to 0, 1, ..., v_3 powers of z (evaluated at some z_i). If the diagram also has v_4 four-vertices, the number of propagators along the line is $v_3 + v_4 - 1$.

The sum over all possible cuts for the particular diagram can then be written as

$$\begin{aligned} &\sum_{j=1}^{v_3+v_4-1} H(m_1 \dots m_k; j) = \\ &= \sum_{j=1}^{v_3+v_4-1} \left[(H_0 + zH_1 + z^2 H_2 + \dots H_{v_3} z^{v_3}) \prod_{q \neq j} \frac{1}{\hat{p}^2_{1q}} \right]_{z=z_j} \frac{1}{p^2_{1j}} \end{aligned} \tag{4.171}$$

or

$$\begin{aligned} \sum_{j=1}^{v_3+v_4-1} H(m_1 \dots m_k; j) \;=\; & H_0 \sum_{j=1}^{v_3+v_4-1} \left[\frac{1}{\prod_{q \neq j} \hat{p}^2_{1q}} \right]_{z=z_j} \frac{1}{p^2_{1j}} \\ & + H_1 \sum_{j=1}^{v_3+v_4-1} \left[\frac{z}{\prod_{q \neq j} \hat{p}^2_{1q}} \right]_{z=z_j} \frac{1}{p^2_{1j}} \\ & + H_2 \sum_{j=1}^{v_3+v_4-1} \left[\frac{z^2}{\prod_{q \neq j} \hat{p}^2_{1q}} \right]_{z=z_j} \frac{1}{p^2_{1j}} \\ & + \dots \\ & + H_{v_3} \sum_{j=1}^{v_3+v_4-1} \left[\frac{z^{v_3}}{\prod_{q \neq j} \hat{p}^2_{1q}} \right]_{z=z_j} \frac{1}{p^2_{1j}} \end{aligned} \tag{4.172}$$

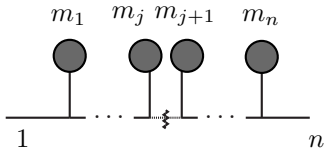
Due to the set of identities eq.(4.136), all terms involving z^λ with $0 < \lambda < v_3 + v_4 - 1$ vanish identically.

Moreover it is easy to see that the z-independent term involving H_0 will give the corresponding Feynman diagram with the help of the identity eq.(4.134).

Let us further distinguish among three cases: diagrams with no four-vertex on the line ($v_3 = k+1$), with one four-vertex ($v_3 = k$) and with more than one four-vertex ($v_3 < k$). In the third case the right hand side of eq.(4.172) reduces to the z-independent term corresponding to the Feynman diagram. All terms involving z vanish. In other words the sum over all possible cuts of the hatted diagrams with two or more four-vertices on the line, is equal to the corresponding Feynman diagrams.

In the first two cases the sum of all hatted diagrams over all possible cuts gives the corresponding Feynman diagrams plus a number of terms. We will now show that these terms cancel each other exactly due to the structure of the YM vertices.

Diagrams with no four-vertex on the line have $v_3 = k+1$ three-vertices. They are of the form

$H(m_1, \ldots, m_j, m_{j+1}, \ldots, m_{k+1}; j) =$ 

(4.173)

The highest-order term involves H_{k+1} that vanishes identically: it corresponds to diagrams with crossed white blobs on the line exclusively, hence it consists of contractions of ϵ with itself or ϵ_1, ϵ_n. The next-to-leading order term, H_k is a sum of terms with one QCD vertex and k crossed white-blob vertices.

We have

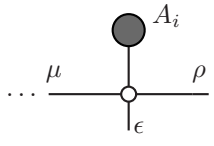
$$\begin{aligned}\sum_{j=1}^{k} H(m_1 \ldots m_{k+1}; j) &= F(m_1 \ldots m_{k+1}) + \\ &\quad + H_k \sum_{j=1}^{k} \left[\frac{z^k}{\prod_{q \neq j} \hat{p}_{1q}^2} \right]_{z=z_j} \frac{1}{p_{1j}^2} \end{aligned} \tag{4.174}$$

and using the kinematical identity eq.(4.139) we get

$$\sum_{j=1}^{k} H(m_1 \ldots m_{k+1}; j) = F(m_1 \ldots m_{k+1}) - H_k \frac{1}{\prod_{j=1}^{k} 2\epsilon \cdot p_{1j}} \tag{4.175}$$

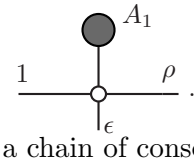
Let us denote $H_{k,r}$ the term where the QCD vertex is coming from the r'th hatted vertex.

$H_{k,r}$ is the numerator of a diagram containing blobs connected with the main line by white-blob vertices $\epsilon^c M_{\mu\nu\rho c}$ everywhere except the QCD r'th vertex. A generic piece on the line will contribute by



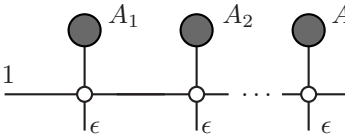
$$\begin{aligned} \equiv \quad & M(A_i)_{\mu\rho} = -J^{\nu}_{A_i} \epsilon^c (2 g_{c\nu} g_{\mu\rho} - g_{c\mu} g_{\nu\rho} - g_{c\rho} g_{\mu\nu}) \\ = \quad & -(2(J_{A_i} \cdot \epsilon) g_{\mu\rho} - J_{A_i,\mu} \epsilon_\rho - J_{A_i,\rho} \epsilon_\mu) \end{aligned} \tag{4.176}$$

where $J^{\nu}_{A_i}$ is the current (subamplitude) coming from the i'th blob. The first such white-blob vertex, when contracted with ϵ_1^μ, gives



$$\ldots = \epsilon_1^\mu M(A_1)_{\mu\rho} = -(\epsilon_1 \cdot J_{A_1}) \epsilon_\rho \tag{4.177}$$

It is easy to see that a chain of consecutive white-blob vertices gives



$$= (-1)^\rho (\epsilon_1 \cdot J_{A_1})(\epsilon \cdot J_{A_2}) \ldots (\epsilon \cdot J_{A_s}) \epsilon_\rho \tag{4.178}$$

Since $H_{k,r}$ has k such vertices we get

$$\begin{aligned} H_{k,r} &= (-1)^k \prod_{j\neq r} (\epsilon\cdot J_{A_j}) \epsilon_1^\mu J_{A_r}^\nu \epsilon_n^\rho \times \\ &\quad \times (g_{\mu\nu}(p_{1,r-1}-p_r)_\rho + g_{\nu\rho}(p_r+p_{1,r})_\mu + g_{\rho\mu}(-p_{1,r}-p_{1,r-1})_\nu) \\ &= (-1)^k \prod_{q\neq r} (\epsilon\cdot J_{A_q})(\epsilon_n\cdot J_{A_r})((p_{1,r-1}+p_{1,r})\cdot\epsilon_1) \end{aligned} \tag{4.179}$$

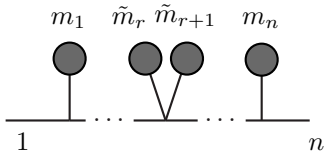
where repeated use was made of our ability to interchange ϵ_1, ϵ_n and ϵ. Note that for the boundary terms, $r = 1$ and $r = k+1$, the above formula must be understood with the definitions $p_{1,0} = p_{1,k+2} = 0$.

Summing over r we regain H_k, and

$$\begin{aligned} \sum_{j=1}^k H(m_1 \dots m_{k+1}; j) &= F(m_1 \dots m_{k+1}) + \\ &\quad + (-1)^{k+1} \frac{1}{\prod_j 2\epsilon\cdot p_{1j}} \sum_r (\epsilon_1\cdot J_{A_1}) \dots (\epsilon_n\cdot J_{A_n})[(p_{1,r-1}+p_{1,r})\cdot\epsilon] \\ &= F(m_1 \dots m_k) + \\ &\quad + (-1)^{k+1} (\epsilon_1\cdot J_{A_1}) \dots (\epsilon_n\cdot J_{A_n}) \sum_r \frac{2p_{1,r}\cdot\epsilon}{(2\epsilon\cdot p_{11})(2\epsilon\cdot p_{12})\dots(2\epsilon\cdot p_{1k})} \\ &= F(m_1 \dots m_k) + \\ &\quad (-1)^{k+1} (\epsilon_1\cdot J_{A_1}) \dots (\epsilon_n\cdot J_{A_n}) \sum_r \frac{1}{\prod_{q\neq r}(2\epsilon\cdot p_{1q})} \end{aligned} \tag{4.180}$$

Let us consider now a diagram that could occur from the ones above by contracting a propagator, thereby merging two of the three-vertices, say the r'th and the $r+1$'th in one four-vertex. Such a diagram has one four-vertex on the line and one propagator less, i.e., it has k vertices of which $k-1$ three-vertices and one four-vertex, as well as $k-1$ propagators.

$H(m_1, \dots, \tilde{m}_r, \tilde{m}_{r+1}, \dots, m_{k+1}; j) =$



(4.181)

The sum over all cuts of such hatted diagrams is

$$\begin{aligned} \sum_{\substack{j=1 \\ j\neq r}}^k H_r(m_1, \dots, \tilde{m}_r, \tilde{m}_r, \dots, m_k; j) &= \\ &= F(m_1, \dots, \tilde{m}_r, \tilde{m}_r, \dots, m_k) + \\ &\quad + H_{k,r} \sum_{\substack{j=1 \\ j\neq r}}^k \left[\frac{z^{k-1}}{\prod_{q\neq j,r} \hat{p}_{1q}^2} \right]_{z=z_j} \frac{1}{p_{1j}^2} \end{aligned} \tag{4.182}$$

where r denotes the position of the four-vertex. As before, all contributions involving z^λ for $0 < \lambda < k-1$ vanish due to eq.(4.136). Using the identity eq.(4.139) we can perform the j sum over cuts and get

$$\begin{aligned} \sum_{\substack{j=1 \\ j\neq r}}^k H_r(m_1, \dots, \tilde{m}_r, \tilde{m}_r, \dots, m_k; j) &= \\ &= F(m_1, \dots, \tilde{m}_r, \tilde{m}_r, \dots, m_k) - \tilde{H}_{k-1,r} \frac{1}{\prod_{q\neq r} 2\epsilon\cdot p_{1q}} \end{aligned} \tag{4.183}$$

Here $\tilde{H}_{k-1,r}$ stands for the vertex contribution of a diagram with one four-vertex and $k-1$ white-blob three-vertices. It is not difficult to see that this is equal to

$$\tilde{H}_{k-1,r} = (-1)^{k-1}(\epsilon_1 \cdot J_{A_1})(\epsilon \cdot J_{A_2}) \dots (\epsilon_n \cdot J_{A_k}) \tag{4.184}$$

Summing up over all possible positions of the four-vertex we get

$$\begin{aligned} \sum_{r=1}^{k} \sum_{\substack{j=1 \\ j \neq r}}^{k} H_r(m_1 \dots m_k; j) = F(m_1, \dots, \tilde{m}_r, \tilde{m}_r, \dots, m_k) + \\ + (-1)^k (\epsilon_1 \cdot J_{A_1}) \dots (\epsilon_n \cdot J_{A_n}) \sum_r \frac{1}{\prod_{q \neq r} 2\epsilon \cdot p_{1q}} \end{aligned} \tag{4.185}$$

which exactly cancels the last term of equation eq.(4.180), thus completing the proof that the sum over all cuts of all possible hatted diagrams is equal to the sum of Feynman diagrams.

4.5.9 Remarks on generality

We have seen in the previous sections how the BCFW decomposition is related to Feynman diagrams. In fact, when working in the particular gauge that we have chosen, every Feynman diagram is broken in pieces ('hatted' graphs) each of which contains a cut in one of the propagators along the line connecting the two 'special' legs. Then, hatted graphs with the same cut are regrouped together in one decomposition, and the sum of decompositions gives back the whole color amplitude.

In the process of breaking up the Feynman graphs in hatted graphs, some extra terms are produced: those that correspond to the $z \to \infty$ limit of the hatted graph. Thanks to the particular structure of the YM vertices, demonstrated in eq.(4.159), these extra terms cancel exactly among graphs with no four-vertices and graphs with one four-vertex and one propagator less. In the lowest case of one propagator along the line, the diagram with two three vertices gives an extra term that is exactly equal to the diagram with no propagator and one four-vertex. The latter is a diagram which, having no propagator along the line, didn't fit in an obvious way in any BCFW decomposition.

The absence of a Yang-Mills structure in a scalar theory is what makes a decomposition along the terms of BCFW cumbersome. In that case one would have to accommodate the diagram in eq.(4.144 by adding an ad hoc term in the recurrence relation which would therefore be less elegant.

The gauge in which we are working explicitly eliminates the z-dependence from the polarization vectors. In other gauges the polarization vectors ϵ_1^+ and ϵ_n^- would be affected by the shift. This would complicate significantly the algebra, as further poles related to vanishing denominators of these polarization vectors will come to play.

If the identity eq.(4.144) doesn't hold, for example, contributions from the class of diagrams in the left hand side of eq.(4.144) would have to be canceled by contributions from other classes. The appealing cancelations within subclasses of diagrams with a particular partitioning of the external legs would be lost. Keeping track of the terms in fragmenting and regrouping Feynman diagrams into BCFW decompositions would be much harder. Still one could prove the BCFW decomposition by performing the pole expansion on any Feynman diagram minus its $z \to \infty$ limit but the algebra would be particularly cumbersome.

It is becoming increasingly clear that the BCFW decomposition is a rearrangement of terms of Feynman diagrams within a color-ordered gluonic amplitude. A similar impression will presumably occur from an approach of amplitudes with fermionic lines. Under the light of the diagrammatic proof presented here it is seen that the mechanism behind any apparent cancelations in the BCFW recursion formula does not (unfortunately) divulge a deeper principle or

even a contingent effect in gluonic amplitudes, but results directly from the structure of the Yang-Mills vertices and the cancelations induced by gauge invariance.

From the point of view of computational complexity it was very early pointed out (and shown explicitly in [21]) that the BCFW formula would be of little help. As an analytical tool, however, it has it merits, and has already offered closed expressions for previously unmanageable helicity amplitudes in tree- and even one-loop level[16]. The analysis presented here will, hopefully, contribute[17] to de-mystifying the method and embedding it in the framework of ordinary field theory.

[16] See [9] for recent one-loop developments

[17] Along with other, similar approaches, see [66].

Appendix A

The construction of Niederreiter sequences

A.1 Introduction

As explained in chapter 3, Quasi-Monte Carlo is a numerical integration method that uses point-sets constructed to be more uniform than random ones. These point-sets cover smoothly the integration region avoiding the kind of clustering that is typical for pseudo-random point-sets. Uniformity is measured by a variety of functions $D(X)$, defined on the point-sets X, collectively called "discrepancies" [1]. Good quality quasi point sequences purport to achieve relatively (to the truly random sets) small values for any discrepancy.

Proving that an (in principle) infinite such sequence (or any of its subsequences) has a particular value for some discrepancy is highly non-trivial. On the other hand, constructing sequences with nice symmetry properties that allow the prediction of a strict asymptotic upper bound for the values of a discrepancy is plausible, and has been achieved by the work of a number of mathematicians, among which van der Corput, Halton, Sobol, Faure and Niederreiter.

Hidden between the lines, here, is the demand that our (finite) point-sequence should have a low discrepancy for any number of points: we don't know in advance how many points will be used in a particular calculation[2]. The art of generating smooth point-sets relies heavily on decisions on the **order** in which the points are generated. The degree in which this demand complicates things can be demonstrated by a simple one-dimensional example: if one is asked to find the point-set with N_0 one-dimensional points having the lowest discrepancy, one can immediately reply that the desired point-set has its points equidistributed along the real axis in the interval from 0 to 1. It is an entirely different matter to decide in which order these points should be generated so that the discrepancy remains as low as possible for any subset of this point-set that might be used in an actual calculation [3].

In this appendix we will see how are the Niederreiter sequences constructed. The Niederreiter sequences are chosen because their asymptotic upper bounds are the lowest that have been achieved presently. The approach, here, will be a constructive one, and no attempt will be made to prove any of the theorems by Niederreiter that support the low discrepancy claims. We will demonstrate explicitly, though, the particularly low values of a diaphony enjoyed by these point-sets.

A.2 Discrepancies and bounds

The most popular discrepancy in the mathematical literature is the so-called star discrepancy $D^*(X)$ defined by

$$D^*(X) = sup\,|A_j(N) - NV_j(N)| \tag{A.1}$$

where N is the number of points in the point-sets X, $A_j(N)$ is the number of points that lie in the subinterval j, $V_j(N)$ is the volume of that subinterval, and the supremum extends over all subintervals

$$J = \prod_{k=1}^{d} [0, u_k) \tag{A.2}$$

for any u_k. The supremum over an infinite amount of subintervals seems impossible to calculate in practice, but things are better than they appear, once we realize that it suffices to calculate $A_j - NV_j$ in the N intervals that are defined by the points themselves (including the defining point).

[1] See section 3.4.1 for a definition of a special kind of discrepancy, named 'diaphony', which is used throughout this thesis.

[2] Moreover, we certainly don't want to alter our point-generating algorithm depending on the number of points used.

[3] The solution is given by the Van der Corput sequence, if N_0 is a power of a small integer.

It is believed, but not proven[4], that the smallest possible bound for D^* is

$$D^*(X) \leq C_d \frac{(logN)^s}{N} + \frac{1}{N} O((logN)^{s-1}) \tag{A.3}$$

where s is the number of dimensions of the point-set. Niederreiter has proposed a sequence that yields the smallest values currently known for C_s. We will give a table with values of C_s later on, once the construction algorithm of the sequence is exposed.

These upper bounds are supposed to approach asymptotically the actual star discrepancy of the sequence, but this doesn't exclude a better behavior in all practical regimes with finite number of points.

Other discrepancies are also available. An extensive discussion of diaphonies has been made in sec.3.4.1.

Discrepancies are usually normalized such that the average value they take on a truly random point-sets is equal to 1. With 0 being their absolute minimum (they are positive definite), a quasi point-sets should perform somewhere in $[0, 1)$, and preferably appreciably closer to 0.

A.3 (t,m,s)-nets

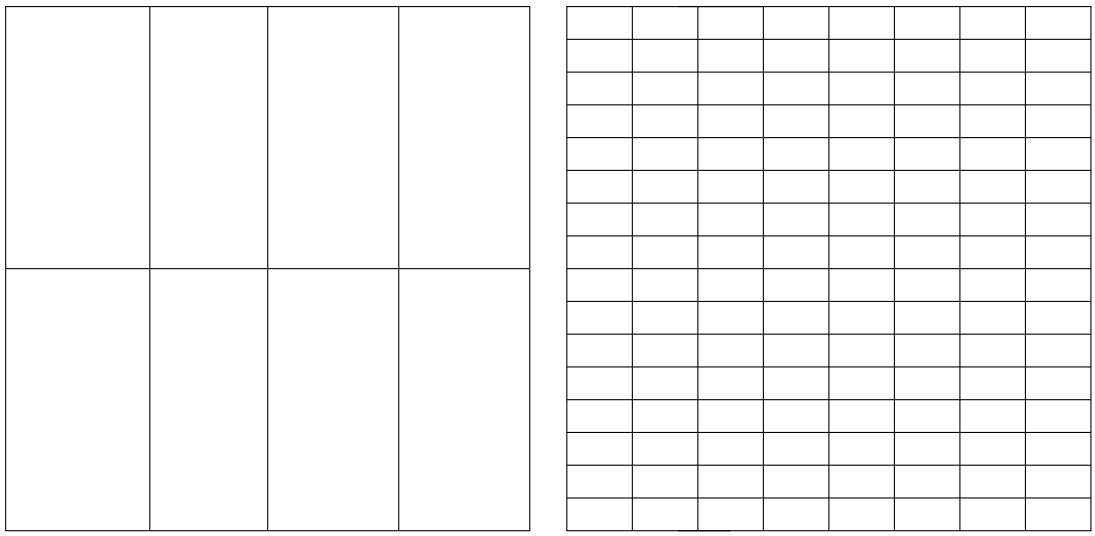


Figure A.1: 2-dimensional elementary intervals with $w_1 = 2$, $w_2 = 1$ (left) and $w_1 = 3$, $w_2 = 4$ (right).

The construction of Niederreiter aims at minimizing the star discrepancy. Hence the quantities $A_j - NV_j$ should be as small as possible in as many as possible different subintervals j. The definition of Elementary Intervals is helpful here. An elementary interval in base b is one of the form

$$E = \prod_{i=1}^{s} [\frac{a_i}{b^{w_i}}, \frac{a_i + 1}{b^{w_i}}) \tag{A.4}$$

with integers $w_i \geq 0$ and $0 \leq a_i < b^{w_i}$. The choice $w_i = 0$ means all $a_i = 0$ and this gives only one interval, the hypercube $[0, 1)$ itself. The choice $w_1 = 1, w_{i\neq 1} = 0$ has $a_1 = 0, 1, .., b-1$ $a_{i\neq 1} = 0$ which gives all the subintervals obtained by cutting the hypercube in b equal pieces along the first dimension. The choice $w_1 = 1, w_2 = 1, w_{i\neq 1,2} = 0$ has $a_1, 2 = 0, 1, .., b-1$ $a_{i\neq 1,2} = 0$ which gives all the subintervals obtained by cutting the hypercube in b equal pieces along the

[4]It is proven for $s \leq 2$ by W. M. Schmidt [61].

first and the second dimension, and so on (see fig.A.3). The goal in the Niederreiter sequence is to divide the available points equally among the elementary intervals. Naturally, this becomes increasingly difficult as w increases. Ideally, however, the algorithm should provide an infinite sequence that fills equally elementary intervals of increasing w, as the sequence evolves. The notion of (t-s)-nets will state this demand in more precise terms.

(t,s)-nets in base b are s-dimensional point sequences with the property that for every $m > t$ there are b^t points out of every pack of b^m that lie in every elementary interval with volume b^{t-m}. For example, a (0,2)-net in base 2 is a sequence for which one out of every two points lie in each elementary interval of volume 1/2, one out of every four points lies in each interval with volume 1/4, one out of eight points lies in each interval with volume 1/8 etc. A (1,2)-net would have one out of four points in any interval of volume 1/4 but not necessarily[5] one out of two in the intervals of volume 1/2. It is evident that the lower t is the more constrained the sequence is. It is also evident that such a sequence would achieve low discrepancy values. We see here an explicit example of how important is the ordering of the points in a point-sets that purports to keep the discrepancy *a*lways low.

The algorithm of Niederreiter constructs such (t-s)-nets for any dimension and base. In [50] an analysis is also given of which is the optimal base for each dimension.

A.4 The algorithm

We follow closely the version of the algorithm[6] presented in [13], the details of which we recite here for convenience. The construction of a (t-s)-net begins with deciding what the base b will be.

Given the base, the way to generate the n-th point of the sequence is as follows: $n-1$ is written in base b and its representation is inverted. If

$$n - 1 = a_0 + a_1 b^1 + a_2 b^2 + \ldots + a_k b^k \quad 0 \leq a_i < b \tag{A.5}$$

then

$$y_n = a_0 b^{-1} + a_1 b^{-2} + \ldots + a_k b^{-k-1} \tag{A.6}$$

is the 'inverted' point. Using arithmetics in base b

$$y_n = 0.a_0 a_1 a_2 \ldots a_k \qquad base\ \ b \tag{A.7}$$

As an example, if $n = 42$ and the base is 2, we have

$$n - 1 = 41 = 32 + 8 + 1 = 1 + 0 * 2 + 0 * 2^2 + 1 * 2^3 + 0 * 2^4 + 1 * 2^5 \tag{A.8}$$

(or $41 \equiv 101001$ in base 2), hence

$$y_{42} = 1 * \frac{1}{2} + 1 * \frac{1}{2^4} + 1 * \frac{1}{2^6} = 0.578125 \tag{A.9}$$

(or $y_n = 0.100101$ in base 2).

Identifying x_n with y_n would reproduce a one-dimensional Halton - Van der Corput sequence which is also very uniform but has been reported to exhibit some troublesome characteristics in

[5] To see the difference, assume we are in $s = 1$ and we look at the four elementary intervals of volume $\frac{1}{4}$, here denoted by $A_1, \ldots, A_4$. In a $(0,1)$-net placing the first point in one of the A_1, A_2 would force the next point to be in one of the A_3, A_4. One possible arrangement for the first four points would be $\{A_1, A_3, A_2, A_4\}$. In the case of a $(1,1)$-net the above arrangement is, of course, possible. However, an arrangement like A_1, A_2, A_3, A_4 would, now, also be possible since the constraint 'place one out of every two points in each interval of volume $\frac{1}{2}$' doesn't have to be satisfied.

[6] A $c++$ implementation of the Niederreiter algorithm is available at *http://www.ru.nl/imapp/theory/mc_qmc.*

higher dimensions. One can do better by permuting the set of y_n in such a way as to produce (t-s)-nets. The permutation is performed with the help of a representation matrix c_{jr}, the construction of which lies at the heart of the algorithm. This matrix acts on the b-dimensional representation of n (whose components are just a_r) but with multiplication and addition defined within a Galois Field GF_b which will be defined shortly.

Given this matrix one sets

$$d_j = \sum_{r=0}^{k-1} c_{jr} a_r \tag{A.10}$$

where all additions and multiplications are performed in the field GF_b and

$$x_n = \sum_{j=1}^{\infty} d_j b^{-j} \tag{A.11}$$

or in base b

$$x_n = 0.d_1 d_2 \dots d_k \tag{A.12}$$

This can be extended to any dimension, s, by picking s different matrices $c^i_{j,r}$, corresponding to (as will be seen) qualitatively different permutations, and using one for every dimension. The concise formula for the i-th coordinate of the n'th point is

$$x^i_n = \sum_{j=1}^{\infty} b^{-j} \sum_{r=0}^{\infty} c^i_{j,r} a_r(n) \qquad i = 1, \dots, s \tag{A.13}$$

where $a_r(n)$ is the coefficient of b^{-r} in the b-base representation of $n-1$.

A.5 The construction of the permutation matrix c_{jr} in one dimension

The basis b can, in principle, be chosen to be any integer. Niederreiter has shown [50], however, that the optimal bases for any dimension up to $s = 20$ is always a prime or a power of a prime. We will restrict ourselves to these cases. The construction of Niederreiter requires, among other things[7], a commutative ring[8] with multiplicative identity, consisting of b elements. Since b is assumed to be a prime, or a power of a prime, one can always recognize such a ring in the Galois Field[9] of order $b = p^k$, usually denoted as GF_{p^k}.

Given the base b and the field GF_b one can define polynomials in one variable x with coefficients in the field, i.e with coefficients that belong to GF_b and obey the arithmetic operations of the field. If the base is a prime p then the polynomials defined above would be polynomials in x with integer coefficients,

$$a_n x^n + a_{n-1} x^{n-1} + \dots a_1 x^1 + a_0 \tag{A.14}$$

[7]Following [13] we will simplify the original algorithm of [50] and set ψ_r and λ_{ij} (as defined in [50] page 54) to be the identity bijections between the commutative ring M and the set of digits in base b.

[8]A ring is a set equipped with two mathematical operations, commonly denoted as $+$ and $*$, such that (a) addition is associative, and commutative, there is an addition identity element and an additive inverse for every element, and (b) multiplication is associative and distributivity is valid ($a*(b+c) = a*b+a*c$). A commutative ring is a ring where multiplication is also commutative. Note that the existence of a multiplicative identity element or a multiplicative inverse *is* **not** assumed.

[9]A Galois Field is a field with finite number of elements. A field is a ring satisfying the extra properties of (a) commutative multiplication, (b) existence of multiplicative identity and (c) existence of multiplicative inverse for every element. There is exactly one Galois Field of order p^k if p is a prime. If $k = 1$, i.e. if the field is of prime order, then it is (isomorphic to) the set of the first p integers $\{0, 1 \dots, p-1\}$ with addition and multiplication modulo p. If $k > 1$, however, this is *n*ot the case. As an example of the later case, one can see that $\{0, 1, 2, 3\}$ with operations modulo 4 is not a field, since $2 \times 2 = 4 \equiv 0$ which shows that 2 has no multiplicative inverse.

with all $a_i < p$. Addition and multiplication are done modulo p, which means, for example, that, in base 3

$$(x+1)(x+2) = x^2 + 3x + 2 = x^2 + 2 \tag{A.15}$$

since $3 = 0 \; mod \; 3$. Irreducible polynomials are those polynomials that cannot be factored in two others with coefficients in the field. In base 2 the polynomial $x^2 + 1$ is *n*ot irreducible, since

$$(x+1)^2 = x^2 + 2x + 1 = x^2 + 1 \quad mod \; 2 \tag{A.16}$$

whereas the polynomial $x^2 + x + 1$ is irreducible as can be easily checked. Further more, a monic irreducible polynomial is an irreducible polynomial with the coefficient of the higher degree in x equal to 1 (or equal to the unit element of multiplication in more complicated cases). The matrix c_{jr} is in principle an infinite size matrix. In practice we will need the upper left square submatrix c_{jr} for $j, r = 1, \ldots, R$ with R an integer determined by the order of magnitude of the number of points we want to generate.

To construct the matrix c_{jr} one has to pick a monic irreducible polynomial in the ring M, $p(x)$. Denoting the degree of $p(x)$ with e we have the following algorithm:

1. Set $j = 0, q = -1, u = 0$

2. Increment j.
   If $u = e$ or $u = 0$ go to step 3. Otherwise go to step 4.

3. Increment q. Set $u = 0$.
   Define
   $$b(x) = p(x)^{q+1} = x^m - b_{m-1}x^{m-1} - \ldots - b_0. \tag{A.17}$$
   Set
   $$v_0 = v_1 = \ldots = v_{m-2} = 0 \;\; , v_{m-1} = 1 \tag{A.18}$$
   $$v_{t \geq m} = \sum_{k=1}^{m} b_{m-k} v_{t-k} \;\; t = m, m+1, \ldots, R+e-2$$
   where addition and multiplication are performed in the field GF_b.

4. Set $c_{jr} = v_{r+u}$ for $r = 0, 1, \ldots, R-1$.
   Set $u \rightarrow u + 1$.
   If $j < R$ go to step 2. Otherwise stop.

As an example we consider the construction of a 6×6 matrix in the case of base 2. Then the corresponding field, GF_2, is just the set $\{0, 1\}$ and all operations are performed modulo 2. The first monic polynomial is of degree 1, $p(x) = x$. We then have

step 1 $j = 0, q = -1, u = 0$

step 2 $j = 1$. We see that $u = 0$ so we proceed to step 3.

step 3 $q = 0$, so $b(x) = p(x)^{0+1} = x$, hence $m = 1, b_0 = 0$. This gives $v_{1r} = 100000$.

step 4 $c_{1r} = 100000$. We set $u = 1$.

step 2 $j = 2$. We see that $u = 1 = e$ so we proceed to step 3.

step 3 We set $u = 0$. We set $q = 1$, so $b(x) = p(x)^2 = x^2$, hence $m = 2$, $b_0 = 0, b_1 = 0$. This gives $v_{2r} = 010000$

step 4 $c_2r = 010000$. We set $u = 1$.

step 2 ...

We see that the algorithm outputs a matrix $c_{jr} = \delta_{j,r}$, equal to the identity matrix. This is the most trivial case and will produce the Halton point-set along this particular dimension. We proceed with the example of base 2, with a monic polynomial $p(x) = x^2 + x + 1$. Then we have

step 1 $j = 0,\ q = -1,\ u = 0$

step 2 $j = 1$. We see that $u = 0$ so we proceed to step 3.

step 3 $q = 0$, so $b(x) = p(x) = x^2 + x + 1 = x^2 - x - 1$, so $m = 2, b_0 = 1, b_1 = 1$. This gives $v_r = 01010101\ldots$

step 4 $c_{1r} = 010101$. We set $u = 1$

step 3 $j = 2$. We see that $u = 1 < 2$ so we jump to step 4.

step 4 $c_{2r} = 101010$. We set $u = 2$

step 2 $j = 3$. We see that $u = 2 = e$ so we go to step 3.

step 3 We set $u = 0$. We set $q = 1$ so

$$b(x) = p(x)^2 = (x^2+x+1)(x^2+x+1) = x^4+x^3+x^2+x^3+x^2+x+x^2+x+1 = x^4-x^2-1 \tag{A.19}$$

so $m = 4, b_0 = 1, b_1 = 0, b_2 = 1, b_3 = 0$ which gives $v_r = 000101000101\ldots$

step 4 $c_{3r} = 000101$. We set $u = 1$.

step 2 $j = 4$. We see that $u = 1 < 2$ so we jump to step 4.

step 4 $c_{4r} = 001010$. We set $u = 2$

step 2 $j = 5$. We see that $u = 2 = e$ so we proceed to step 3.

step 3 We set $u = 0$. We set $q = 2$ so

$$b(x) = p(x)^3 = (x^4+x^2+1)(x^2+x+1) = x^6+x^5+x^4+x^4+x^3+x^2+x^2+x+1 = x^6-x^5-x^3-x-1 \tag{A.20}$$

so $m = 6, b_0 = 1, b_1 = 1, b_2 = 0, b_3 = 1, b_4 = 0, b_5 = 1$. This gives $v_r = 0000011101110\ldots$

step 4 $c_{5r} = 000001$. We set $u = 1$

step 2 $j = 6$. We see that $u = 1 < e$ so we jump to step 4.

step 4 $c_{6r} = 000011$. We see that $j = 6 = R$ so the algorithm exits.

The net result of the above is a matrix

$$c = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \tag{A.21}$$

which is certainly non-trivial.

A.6 The generalization in more dimensions

The extension of the algorithm to D dimensions is straightforward. For each dimension one picks a monic polynomial $p_i(x)$ and constructs the corresponding matrix $c_{i,jr}$.

A.7 Generating Niederreiter points

The n'th point of the sequence, which has coordinates

$$\vec{x}_n = (x_{n,1}, x_{n,2} \ldots, x_{n,s}) \tag{A.22}$$

is found by

- finding the inverse of $n-1$ in base b

$$n-1 = a_0 + a_1 b^1 + a_2 b^2 + \ldots + a_k b^k \quad 0 \leq a_i < b \tag{A.23}$$

- finding the coefficients

$$d_{i,j} = \sum_{r=0}^{R-1} c_{i,jr} a_r \quad j = 1 \ldots R \tag{A.24}$$

- and setting

$$x_{n,i} = 0.d_{i,1} d_{i,2} \ldots d_{i,R} \tag{A.25}$$

As an example we will use the two matrices that were produced in section A.5, for a sequence in base 2. As we saw there, the choice $p_1(x) = x$ results to $c_{ij} = \delta_{i,j}$, so one gets the Halton sequence. We shall represent the first 2^k points of the one-dimensional Halton sequence by $[h_1, h_2, \ldots, h_N]$, where h_i indicates the order in which the point with coordinate $(i-1)/2^k$ is produced[10]. Then the first 16 points of the Halton sequence appear in the ordering $[1, 9, 5, 13, 3, 11, 7, 15, 2, 10, 6, 14, 4, 12, 8, 16]$.

For the second coordinate we have to use another monic polynomial, and in this case we shall use $p(x) = x^2 + x + 1$. Then acting with the matrix eq.(A.21), we get a permutation of the first 16 points in the Halton sequence resulting in $[1, 6, 16, 11, 2, 5, 15, 12, 3, 8, 14, 9, 4, 7, 13, 10]$. Note that the matrix permutes the first 4 points among themselves and the first 16 points among themselves, but the first 2 and 8 points are mixed with points produced after them.

In general, a monic polynomial of degree k in base b produces a $c-$matrix that permutes packs of b^{km} points for every $m \geq 0$. This is, of course, directly related to the Corollary 1 in [50], which will be presented shortly (see eq.(A.26)).

A.8 Remarks on the algorithm

The algorithm described above provides $(t-s)$-nets for any dimension s and relatively small values of t. Corollary 1 in [50] shows that

$$t = T_b(s) = \sum_{i=1}^{s} (e_i - 1) \quad e_i \equiv deg(p_i(x)). \tag{A.26}$$

In other words, t is equal to the sum of the degrees of the chosen polynomials minus one.

If the monic polynomial $p_i(x)$ is of degree 1 the condition of step 2 is always fulfilled. For each j a new vector $\vec{v}$ is calculated and $c_{jr} = v_r$. This, in turn, means that the matrix c_{jr} is actually upper diagonal, which makes the coordinates x_i of every pack of b^k points a permutation

[10] The equivalent for 4 points would be $[1, 3, 2, 4]$ indicating that the output sequence is $\{0.0, 0.5, 0.25, 0.75\}$.

of the Halton sequence. This checks against the fact that in such a case the net produced is a $(0-s)$-net.

If the base is chosen to be prime there are always b monic irreducible polynomials of degree 1, namely $p_i(x) = x + i - 1$. If moreover $b > s$ one can always choose first degree polynomials, simplifying the algorithm for constructing $c_{i,jr}$. As an example, consider the construction of a 3-dimensional point-sets. We pick $b = 3$ and our polynomials as $p_1(x) = x$, $p_2(x) = x + 1$, $p_3(x) = x + 2$. Then we get the following three c-matrices:

$$c_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad c_2 = \begin{pmatrix} 1 & 2 & 1 & 2 & 1 & 2 \\ 0 & 1 & 1 & 0 & 2 & 2 \\ 0 & 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad c_3 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 0 & 1 & 2 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{A.27}$$

As a side remark, when $p(x) = x + i$ one can find $c_{j,r}$ directly from the following matrix

$$Q = \begin{pmatrix} 1 & -1 & 1 & -1 & 1 & -1 \\ 0 & 1 & -2 & 3 & -4 & 5 \\ 0 & 0 & 1 & -3 & 6 & -10 \\ 0 & 0 & 0 & 1 & -4 & 10 \\ 0 & 0 & 0 & 0 & 1 & -5 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{A.28}$$

since

$$c_{j,k} = Q_{j,k} \times i^{k-j} \ \ mod\ b \tag{A.29}$$

A.9 Optimal bases

Niederreiter in [50] cites for every dimension a base that minimizes the constant C_s in (Eq.A.3). If the point-sequence is a $(t-s)$-net then $C_s = C(t, s, b)$ and

$$C(t, 2, b) = \frac{1}{8} b^t \left(\frac{b-1}{log(b)}\right)^2 \ \ s = 2 \tag{A.30}$$

$$C(t, 3, 2) = \frac{2^t}{24(log(2))^3} \ \ s = 3\ b = 2 \tag{A.31}$$

$$C(t, 4, 2) = \frac{2^t}{64(log(2))^4} \ \ s = 4\ b = 2 \tag{A.32}$$

and in all other cases

$$C(t, s, b) = \frac{1}{s!} b^t \frac{b-1}{2\lfloor b/2 \rfloor} \left(\frac{\lfloor b/2 \rfloor}{log(b)}\right)^s \tag{A.33}$$

We saw in the previous section that t depends on the degrees of the chosen irreducible polynomials. It is evidently optimal to choose the irreducible polynomials of the smallest possible degree thus minimizing $C(t, s, b)$ for given s, b. The question immediately arising is how many irreducible polynomials are there for a given field F_b (base b) of degree n. The answer is given by the formula

$$N_b(n) = \frac{1}{n} \sum_{d|n} \mu\left(\frac{n}{d}\right) b^d \tag{A.34}$$

where $\mu(x)$ is the Möbius function and the sum is over all divisors d of n. The number of irreducible polynomials of degree 1 is b as stated in the previous section. Hence whenever $b > s$ we have $t = 0$.

The optimal pair t, b for every s can be found as follows: if $b \geq s$ then t should be 0 and two cases are of interest. Either b is the smallest even prime power greater than s or the smallest odd power. If $b < s$ one has to find the minimum t for each case separately. The minimizing pair t, b will be found in one of those finite combinations.

For example, $s = 4$ can have $b = 5$ (C(0,4,5)=0.09936), $b = 4$ (C(0,4,4)=0.135) or $b < 4$. In the latter case we have either $b = 2$ or $b = 3$. In the case $b = 2$, since there are two polynomials of degree 1, one of degree 2 and two of degree 3 in base 2, we get $t = 2(1-1)+(2-1)+(3-1) = 3$ (C(3,4,2)=0.5415). In the case $b = 3$ we have three polynomials of degree 1 and three of degree 3 so $t = 1$ (C(1,4,3)=0.0858). The minimizing choice is $b = 3, t = 1$.

It is worth noticing that the discussion on the discrepancy bounds is only meaningful in the asymptotic limit $N \to \infty$. The minimization of C_s indicates better asymptotic behavior and, with a leap of faith, this is perceived as an indication for a better pointsequence in the non-asymptotic regime. Since the function $\frac{log(N)^s}{N}$ blows up in the "intermediate" $N = O(10^5 - 10^{50})$ regime, it is not clear whether the selection of an optimal base for every dimension is of any practical importance. An indicative table with values of s, optimal b, C_s and bounds for $D^*(N)$ with $N = 10^5$ illuminates the issue.

s	b	C_s	bound
3	3	0.1256942444	0.001918105412
4	3	0.08580887382	0.01507561279
5	5	0.02469434978	0.04994891518
6	7	0.01864911954	0.4342826618
7	7	0.004107322124	1.101180763
8	9	0.002992015371	9.235265061
9	9	0.00060521105687	21.50691363
10	11	0.0004281883971	175.1829459
11	11	0.00008116746877	382.3178946
12	13	0.00005604443614	3039.210777
13	13	0.00001008466805	6296.157985
14	13	0.00002190530879	157452.3265
15	17	$0.4422430483\ 10^{-5}$	365970.7788
16	17	$0.7804619608\ 10^{-6}$	743571.6603
17	17	$0.1296323899\ 10^{-6}$	0.142190411010^7
18	19	$0.8471846909\ 10^{-7}$	0.106984429210^8
19	19	$0.1362901436\ 10^{-7}$	0.198149333910^8
20	23	$0.3278000433\ 10^{-7}$	0.548684721510^9

A.10 Implication of choosing non-prime base

If the base is a prime number p, the corresponding Galois Field GF_p is equivalent with the field of the first p integers with addition and multiplication defined modulo p.

If the base chosen is a prime $p$ower $b = p^k$ the corresponding GF_{p^k} is non-trivial. The construction of the multiplication tables in GF_{p^k} can be performed by an algorithm of operations on the smaller Galois Field GF_p:

- label the elements of the field by the first p^k integers: $0, 1, 2, \ldots, p^k - 1$.
- make a correspondence between each element q of GF_{p^k} with a polynomial over GF_p:

$$j \to p_q(x) = a_0 + a_1 x + \ldots + a_{k-1} x^{k-1} \ : \ q = a_0 + a_1 p + \ldots + a_{k-1} p^{k-1} \tag{A.35}$$

  This can be seen as the base-p representation of q.

- find a monic irreducible polynomial of degree k in the field GF_p which will be called the "generating polynomial" $g(x)$. Obviously, the generating polynomial does not correspond to any member of GF_p^k - all polynomials that do are of degree smaller than k.
- the result of adding $q + w$ in GF_{p^k} is found by adding the corresponding polynomials (modulo p) and converting the result back to a new integer:

$$q + w \rightarrow p_q(x) + p_w(x) = p_m(x) \rightarrow m \tag{A.36}$$

- the result of multiplying $q * w$ is found by multiplying the corresponding polynomials, finding the remainder of the division with the generating polynomial, and converting back to an integer:

$$q * w \rightarrow (p_q(x) * p_w(x))\, mod g(x) = p_m(x) \rightarrow m \tag{A.37}$$

As an example let's calculate the tables for base $4 = 2^2$. We write the elements of $GF_4 = 0, 1, 2, 3$ and make the correspondences $0 \rightarrow 0$, $1 \rightarrow 1$, $2 \rightarrow x$, $3 \rightarrow x + 1$. Then, for example,

$$\begin{aligned} 1 + 3 &\rightarrow 1 + (x + 1) = x + 2 \equiv x \rightarrow 2 \\ 3 + 3 &\rightarrow (x + 1) + (x + 1) = 2x + 2 \equiv 0 \rightarrow 0 \end{aligned}$$

The generator is an irreducible polynomial of degree 2 in GF_2, so we can choose $x^2 + x + 1$. Then, for example

$$\begin{aligned} 2 * 3 \quad \rightarrow \quad & x(x + 1)\, mod\, (x^2 + x + 1) = (x^2 + x)\, mod\, (x^2 + x + 1) = \\ & = (x^2 + x + 1)\, mod\, (x^2 + x + 1)\ - (1)\, mod\, (x^2 + x + 1) \\ & \equiv -1 \equiv 1 \rightarrow 1 \end{aligned}$$

Once the addition and multiplication tables are found monic irreducible polynomials in GF_{p^k} can be calculated by a sieve method and using the tables to perform any multiplication or addition of coefficients.

It should be clear that opting for a non-prime base is complicating the initialization of the Niederreiter algorithm and, what is more important, the generation of the next s-dimensional point in the sequence. The only argument supporting such a choice is the minimization of the upper bounds explained in section A.9. In practice it is not clear at all that this choice gives any advantage over choosing, for example, the lowest prime that exceeds s as a base (which, incidentally, results in the Faure sequence [31]). One could, as a matter of fact, go further and adopt base 2 for any dimensionality (which would be equivalent to the Sobol sequence [63]). This is the choice recommended by Bratley and Fox in [13], where they argue that the generation of points in base 2 is faster (since one can then take advantage of binary operations that are built-in in computers) and gives better results for their test integrals whenever the dimensionality exceeds[11] 7.

A direct measure of the relative uniformity between such choices would of course be their relative diaphony. In fig. A.10 and fig. A.10 we see the diaphony for the Niederreiter sequence in $b = 2$ and in optimal base, for the cases[12] $s = 3, 4, 6, 8$. As expected, the $b = 2$ algorithm wins over the optimal base algorithm for dimensionality greater than 7. This might explain the results in [13] section 5.

[11] The authors have experimented with a number of test functions similar to ours and have established that whenever $s > 8$ their estimates approached the expected values for the integrals faster when $b = 2$ than when b is the 'optimal' choice - see the discussion in section 5 of [13] and the tables therein.

[12] When $s = 2$, $b = 2$ is the optimal choice.

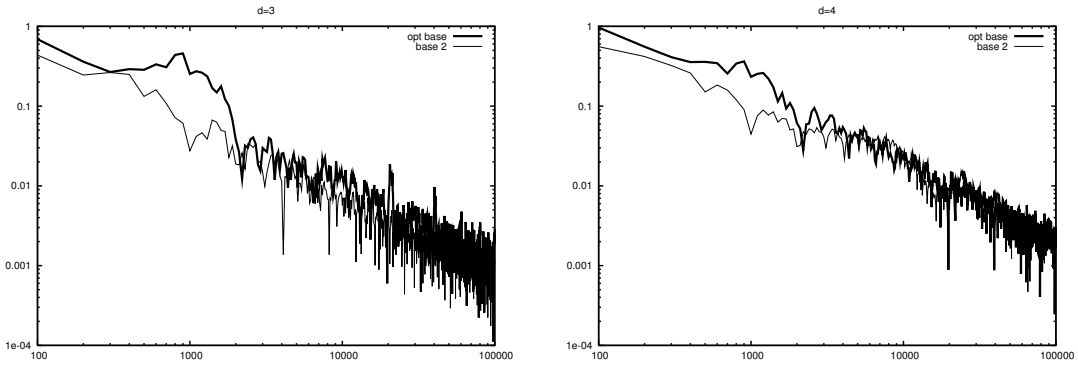


Figure A.2: The diaphony for the Niederreiter sequence in $b = 2$ and in optimal base, for the cases $s = 3, 4$.

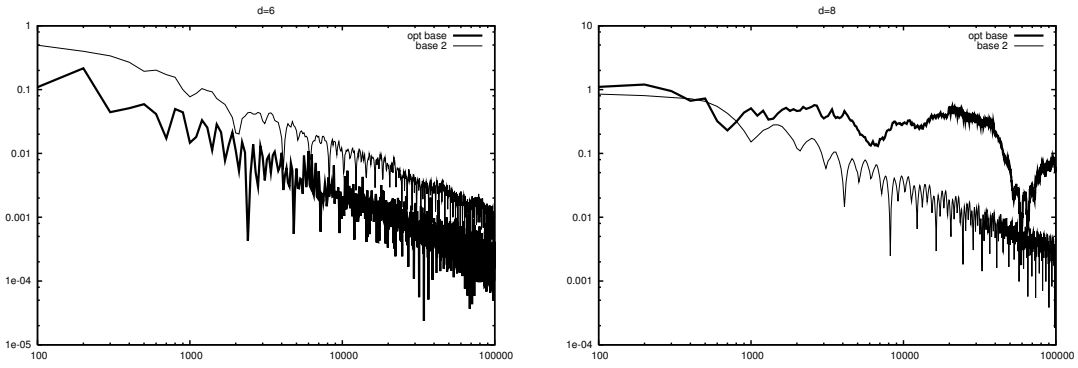


Figure A.3: The diaphony for the Niederreiter sequence in $b = 2$ and in optimal base, for the cases $s = 6, 8$.

Appendix B

Estimators by diagrammatics

B.1 Diagrammatics for Quasi-Monte Carlo and Monte Carlo

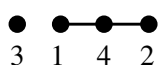
Our strategy for obtaining the form of the estimators is best described by an example. Consider the triple sum

$$S_{p_1} S_{p_2} S_{p_3} \equiv \sum_{i,j,k=1}^{N} f_i^{p_1} f_j^{p_j} f_k^{p_3} \quad . \tag{B.1}$$

In our approach we need to compute the expectation value of this object including the first sub-leading order in $1/N$. It is given by

$$\begin{aligned} \langle S_{p_1} S_{p_2} S_{p_3} \rangle &= N^{\underline{3}} \int f_i^{p_1} f_j^{p_j} f_k^{p_3} \left(1 - \frac{1}{N} \left(F_2(i,j) + F_2(i,k) + F_2(j,k) \right) \right) \\ &\quad + N^{\underline{2}} \int \left(f_i^{p_1+p_2} f_k^{p_3} + f_i^{p_1+p_3} f_j^{p_2} + f_i^{p_1} f_j^{p_2+p_3} \right) + \mathcal{O}(N) \\ &\approx N^3 \int f_i^{p_1} f_j^{p_j} f_k^{p_3} - N^2 \int f_i^{p_1} f_j^{p_j} f_k^{p_3} \left(\alpha_{ij} + \alpha_{ik} + \alpha_{jk} \right) \\ &\quad + N^2 \int \left(f_i^{p_1+p_2} f_k^{p_3} + f_i^{p_1+p_3} f_j^{p_2} + f_i^{p_1} f_j^{p_2+p_3} \right) \quad , \end{aligned} \tag{B.2}$$

with implied integration over the subscripts. The sub-leading terms in the expectation value are, therefore, obtained by either connecting any two of the summands in the multiple sum Ω with a factor $-\alpha$, or by contracting them. Now, any estimator E consists of a linear combination of terms like the above. Its variance, $\langle E^2 \rangle - \langle E \rangle^2$, contains both leading and sub-leading terms. The leading terms, however, cancel completely, and so do the sub-leading terms coming from a connection/contraction *inside* one of the factors E. We arrive at the following diagrammatic prescription. A sum of powers of f will be represented by a labeled dot, and a connection (including the $-\alpha$) by a link between dots. For example,



$$= \sum_{i,j,k,l=1}^{N} f_i^3 f_j f_k^4 f_l^2 \alpha_{jk} \alpha kl \quad . \tag{B.3}$$

Now, suppose that the estimator E is given as a linear combination of *connected* diagrams. The estimator of its variance is the given by the *connected sub-leading* diagrams that can be obtained from $E \times E$. The factors $1/N$ can be added in a straightforward manner: each sum with p different summing indices carries a factor N^{-p}, and there is an additional overall factor N^{1-2^k} in E_{2^k}.

B.2 Estimators for Quasi-Monte Carlo

We apply the above considerations to the first estimators $E^{(q)}_{1,2,4}$ for Quasi-Monte Carlo. Squaring and constructing the connected sub-leading diagrams, we find

$$
\begin{aligned}
E^{(q)}_1 &= \bullet_{1} \\
E^{(q)}_2 &= \bullet_{1}\!\!-\!\!\bullet_{1} + \bullet_{2} \\
E^{(q)}_4 &= 4\,\bullet_{1}\!\!-\!\!\bullet_{1}\!\!-\!\!\bullet_{1}\!\!-\!\!\bullet_{1} + 4\,\bullet_{1}\!\!-\!\!\bullet_{2}\!\!-\!\!\bullet_{1} + 4\,\bullet_{2}\!\!-\!\!\bullet_{1}\!\!-\!\!\bullet_{1} + 4\,\bullet_{3}\!\!-\!\!\bullet_{1} \\
&\quad + \bullet_{2}\!\!-\!\!\bullet_{2} + \bullet_{4} \quad .
\end{aligned}
\tag{B.4}
$$

Upon insertion of the correct factors of $1/N$, we arrive precisely at the estimators $E^{(q)}_{1,2,4}$ given in this paper. The construction of $E^{(q)}_8$ is straightforward: at that order, tree diagrams with branches develop. It may be worth noting that in this diagrammatic approach it becomes immediately clear that no diagrams with loops (that is, occurrences of α_{jj}, or $\alpha_{ij}\alpha_{ji}$, or $\alpha_{ij}\alpha_{jk}\alpha_{ki}$, and so on) are possible to this order in $1/N$.

B.3 Estimators for Monte Carlo

The MC estimators are of course precisely those of Quasi-Monte Carlo, with the replacement $\alpha_{ij} \to 1$. This means that the topology of the tree diagrams becomes irrelevant, and we can feasibly go up to E_{16}. We find

$$
E_K = \frac{1}{N^{2K-1}} \sum_{s=0}^{K-1} E_{K,s} N^s \quad , \quad K = 1, 2, 4, 8, 16 \ , \tag{B.5}
$$

where the coefficients of the various powers of N are given by

$$
E_{1,0} = S_1 \ , \tag{B.6}
$$

$$
\begin{aligned}
E_{2,0} &= -S_1^2 \ , \\
E_{2,1} &= S_2 \ ,
\end{aligned}
\tag{B.7}
$$

$$
\begin{aligned}
E_{4,0} &= -4S_1^4 \ , \\
E_{4,1} &= 8S_1^2 S_2 \ , \\
E_{4,2} &= -S_2^2 - 4S_1 S_3 \ , \\
E_{4,3} &= S_4 \ ,
\end{aligned}
\tag{B.8}
$$

$$
\begin{aligned}
E_{8,0} &= -256S_1^8 \ , \\
E_{8,1} &= 1024S_1^6 S_2 \ , \\
E_{8,2} &= -1152S_1^4 S_2^2 - 512S_1^5 S_3 \ , \\
E_{8,3} &= 352S_1^2 S_2^3 + 832S_1^3 S_2 S_3 + 224S_1^4 S_4 \ , \\
E_{8,4} &= -4S_2^4 - 224S_1 S_2^2 S_3 - 128S_1^2 S_3^2 - 208S_1^2 S_2 S_4 - 96S_1^3 S_5 \ , \\
E_{8,5} &= 32S_2 S_3^2 + 8S_2^2 S_4 + 48S_1 S_3 S_4 + 48S_1 S_2 S_5 + 32S_1^2 S_6 \ , \\
E_{8,6} &= -S_4^2 - 8S_3 S_5 - 4S_2 S_6 - 8S_1 S_7 \ , \\
E_{8,7} &= S_8 \ ,
\end{aligned}
\tag{B.9}
$$

$$
\begin{aligned}
E_{16,0} &= -4194304S_1^{16} \ , \\
E_{16,1} &= 33554432S_1^{14} S_2 \ , \\
E_{16,2} &= -104857600S_1^{12} S_2^2 - 16777216S_1^{13} S_3 \ ,
\end{aligned}
$$

$$
\begin{aligned}
E_{16,3} &= 162922496S_1^{10}S_2^3 + 93585408S_1^{11}S_2S_3 + 7733248S_1^{12}S_4 \ , \\
E_{16,4} &= -132579328S_1^8S_2^4 - 189530112S_1^9S_2^2S_3 - 20185088S_1^{10}S_3^2 \\
&\quad -37552128S_1^{10}S_2S_4 - 3538944S_1^{11}S_5 \ , \\
E_{16,5} &= 54444032S_1^6S_2^5 + 172064768S_1^7S_2^3S_3 + 69861376S_1^8S_2S_3^2 + 63553536S_1^8S_2^2S_4 \\
&\quad +15532032S_1^9S_3S_4 + 14942208S_1^9S_2S_5 + 1507328S_1^{10}S_6 \ , \\
E_{16,6} &= -9806848S_1^4S_2^6 - 69660672S_1^5S_2^4S_3 - 77729792S_1^6S_2^2S_3^2 - 45197312S_1^6S_2^3S_4 \\
&\quad -43855872S_1^7S_2S_3S_4 - 5931008S_1^8S_3S_5 - 21135360S_1^7S_2^2S_5 \\
&\quad -622592S_1^9S_7 - 5357568S_1^8S_2S_6 - 8060928S_1^7S_3^3 - 2802688S_1^8S_4^2 \ , \\
E_{16,7} &= 551936S_1^2S_2^7 + 14180352S_1^5S_2S_3^3 + 10500096S_1^3S_2^5S_3 + 32006144S_1^4S_2^3S_3^2 \\
&\quad +12816384S_1^4S_2^4S_4 + 6193152S_1^6S_2S_4^2 + 36679680S_1^5S_2^2S_3S_4 \\
&\quad +7016448S_1^6S_3^2S_4 + 13725696S_1^6S_2S_3S_5 + 11722752S_1^5S_2^3S_5 \\
&\quad +2007040S_1^7S_4S_5 + 6072320S_1^6S_2^2S_6 + 1994752S_1^7S_3S_6 \\
&\quad +250880S_1^8S_8 + 1798144S_1^7S_2S_7 \ , \\
E_{16,8} &= -256S_2^8 - 6438912S_1^3S_2^2S_3^3 - 3819520S_1^2S_2^4S_3^2 - 366592S_1S_2^6S_3 \\
&\quad -1046016S_1^2S_2^5S_4 - 3568128S_1^4S_2^2S_4^2 - 8730624S_1^4S_2S_3^2S_4 \\
&\quad -2233344S_1^3S_2^4S_5 - 1807360S_1^5S_3S_4^2 - 9879552S_1^3S_2^3S_3S_4 \\
&\quad -8638464S_1^4S_2^2S_3S_5 - 2035712S_1^5S_3^2S_5 - 342016S_1^6S_5^2 - 2471936S_1^4S_2^3S_6 \\
&\quad -3492864S_1^5S_2S_4S_5 - 1542144S_1^5S_2^2S_7 - 570368S_1^6S_2S_8 - 618496S_1^6S_3S_7 \\
&\quad -607232S_1^6S_4S_6 - 851968S_1^4S_3^4 - 96256S_1^7S_9 - 3602432S_1^5S_2S_3S_6 \ , \\
E_{16,9} &= 542208S_1S_2^4S_3S_4 + 2359296S_1^2S_2^2S_3^2S_4 + 1404160S_1^3S_2S_3S_4^2 \\
&\quad +1608704S_1^3S_2^2S_3S_6 + 514432S_1^2S_2^3S_4^2 + 924672S_1^4S_3S_4S_5 \\
&\quad +765952S_1^4S_2S_4S_6 + 552960S_1^2S_2S_3^4 + 1405952S_1^2S_2^3S_3S_5 \\
&\quad +1814528S_1^3S_2S_3^2S_5 + 540672S_1S_2^3S_3^3 + 1394688S_1^3S_2^2S_4S_5 \\
&\quad +262656S_1^2S_2^4S_6 + 808960S_1^4S_2S_3S_7 + 90624S_1S_2^5S_5 + 575488S_1^3S_3^3S_4 \\
&\quad +451584S_1^4S_2S_5^2 + 191488S_1^5S_5S_6 + 475136S_1^4S_3^2S_6 + 164864S_1^5S_4S_7 \\
&\quad +422912S_1^3S_2^3S_7 + 179200S_1^5S_3S_8 + 343296S_1^4S_2^2S_8 + 167936S_1^5S_2S_9 \\
&\quad +1024S_2^6S_4 + 33792S_1^6S_{10} + 133760S_1^4S_4^3 + 60416S_2^5S_3^2 \ , \\
E_{16,10} &= -174336S_1S_2^2S_3S_4^2 - 191488S_1S_2S_3^3S_4 - 49920S_1^2S_2S_4^3 - 120832S_1S_2^3S_3S_6 \\
&\quad -290304S_1^2S_2S_3^2S_6 - 183936S_1^2S_2^2S_4S_6 - 242688S_1S_2^2S_3^2S_5 - 99328S_1S_2^3S_4S_5 \\
&\quad -87040S_1^3S_4^2S_5 - 231936S_1^2S_2^2S_3S_7 - 134400S_1^3S_2S_4S_7 - 153728S_1^3S_2S_3S_8 \\
&\quad -174080S_1^3S_2S_5S_6 - 64512S_2^3S_3^2S_4 - 105216S_1^2S_3^2S_4^2 - 172288S_1^3S_3S_4S_6 \\
&\quad -29184S_2^4S_3S_5 - 100352S_1^2S_3^3S_5 - 105472S_1^3S_3S_5^2 - 111360S_1^2S_2^2S_5^2 \\
&\quad -90112S_1^3S_3^2S_7 - 22528S_1S_2^4S_7 - 40000S_1^4S_4S_8 - 47104S_1^4S_5S_7 \\
&\quad -68608S_1^3S_2^2S_9 - 49088S_1^2S_2^3S_8 - 47616S_1^4S_3S_9 - 42240S_1^4S_2S_{10} \\
&\quad -10752S_1^5S_{11} - 1152S_2^4S_4^2 - 23552S_2^2S_3^4 - 512S_2^5S_6 - 23296S_1^4S_6^2 \\
&\quad -456960S_1^2S_2S_3S_4S_5 - 16384S_1S_3^5 \ , \\
E_{16,11} &= 3328S_2^3S_5^2 + 20352S_2^2S_3S_4S_5 + 16576S_1S_2S_4^2S_5 + 25088S_1S_3^2S_4S_5 \\
&\quad +27136S_1S_2S_3S_5^2 + 4096S_3^4S_4 + 33024S_1^2S_3S_5S_6 + 26240S_1^2S_2S_3S_9 \\
&\quad +32768S_1S_2S_3^2S_7 + 17536S_1S_2^2S_4S_7 + 5184S_1S_3S_4^3 + 24064S_1^2S_3S_4S_7 \\
&\quad +25856S_1^2S_2S_5S_7 + 19456S_1S_2^2S_3S_8 + 16448S_1^2S_2S_4S_8 + 20224S_1S_2^2S_5S_6 \\
&\quad +11392S_2S_3^2S_4^2 + 13440S_1^2S_2S_6^2 + 11328S_1^2S_4^2S_6 + 10240S_2S_3^3S_5 \\
&\quad +16000S_1^2S_4S_5^2 + 224S_2^4S_8 + 11520S_2^2S_3^2S_6 + 832S_2^3S_4S_6 + 13312S_1S_3^3S_6
\end{aligned}
$$

$$
\begin{aligned}
& +13568S_1^2S_3^2S_8 + 9472S_1^3S_6S_7 + 6912S_2^3S_3S_7 + 9984S_1^3S_5S_8 \\
& +8832S_1^3S_2S_{11} + 10368S_1^3S_3S_{10} + 4224S_1S_2^3S_9 + 8576S_1^3S_4S_9 \\
& +10176S_1^2S_2^2S_{10} + 352S_2^2S_4^3 + 46336S_1S_2S_3S_4S_6 + 3008S_1^4S_{12} \ , \\
E_{16,12} \ = \ & -768S_3^2S_5^2 - 2944S_2S_3S_5S_6 - 1024S_3S_4^2S_5 - 2208S_1S_2S_5S_8 - 1664S_1S_2S_4S_9 \\
& -1216S_2S_4S_5^2 - 3072S_1S_2S_3S_{10} - 2944S_2S_3S_4S_7 - 3584S_1S_3S_5S_7 \\
& -2944S_1S_2S_6S_7 - 2016S_1S_3S_4S_8 - 3008S_1S_4S_5S_6 - 768S_1^2S_7^2 - 1024S_3^3S_7 \\
& -1536S_3^2S_4S_6 - 1472S_1^2S_6S_8 - 224S_2S_4^2S_6 - 1920S_1S_3S_6^2 - 1024S_1S_4^2S_7 \\
& -1472S_2S_3^2S_8 - 1408S_2^2S_5S_7 - 208S_2^2S_4S_8 - 960S_1S_2^2S_{11} - 1792S_1^2S_3S_{11} \\
& -1440S_1^2S_2S_{12} - 1312S_1^2S_4S_{10} - 1792S_1S_3^2S_9 - 1088S_2^2S_3S_9 \\
& -1856S_1^2S_5S_9 - 768S_1S_5^3 - 128S_2^2S_6^2 - 704S_1^3S_{13} - 4S_4^4 - 96S_2^3S_{10} \ , \\
E_{16,13} \ = \ & 128S_2S_7^2 + 128S_5^2S_6 + 32S_4S_6^2 + 160S_3S_5S_8 + 48S_2S_6S_8 + 8S_4^2S_8 \\
& +256S_3S_6S_7 + 192S_4S_5S_7 + 128S_3^2S_{10} + 224S_1S_5S_{10} + 160S_2S_5S_9 \\
& +128S_3S_4S_9 + 192S_1S_6S_9 + 160S_1S_7S_8 + 224S_1S_3S_{12} + 48S_2S_4S_{10} \\
& +32S_2^2S_{12} + 192S_2S_3S_{11} + 128S_1S_4S_{11} + 160S_1S_2S_{13} + 128S_1^2S_{14} \ , \\
E_{16,14} \ = \ & -S_8^2 - 16S_5S_{11} - 16S_7S_9 - 8S_6S_{10} \\
& -4S_4S_{12} - 16S_3S_{13} - 16S_1S_{15} - 8S_2S_{14} \ , \\
E_{16,15} \ = \ & S_{16} \ . \qquad \text{(B.10)}
\end{aligned}
$$

The number of individual terms in each E_K is that of the partitions $\Pi(K)$ of K: $\Pi(1) = 1$, $\Pi(2) = 2$, $\Pi(4) = 5$, $\Pi(8) = 22$, and $\Pi(16) = 231$. Likewise, the number of terms in each $E_{K,s}$ is the partition of K into $(K - s)$ parts. We have not extended our results to the fifth-order error estimator with $K = 32$ and $\Pi(32) = 8349$, since already E_8 and E_{16} are purely academic and we have included them only as an illustration of the method.

B.4 The $O(\frac{1}{N^2})$ contribution to G_p

The second-order contribution to G_p can be found by summing up $O(\frac{1}{N^2})$ terms coming from

1. the pure rings (containing only 2-point vertices).
2. the three graphs contributing to $G_p^{(1,2,3)}$ (containing one 4-vertex, two 3-vertices or two external points).
3. products of a pure ring and one of the three graphs above or two of the graphs above.
4. the new graphs (containing one 6-vertex,one 5-vertex and one 3-vertex, two 4-vertices, one 4-vertex and two. 3-vertices, four 3-vertices, one 3-vertex and three external points, two 3-vertices and two external points or one 4-vertex and two external points).

After a lengthy but straightforward calculation (involving some cancelations) we get

$$
\begin{aligned}
G_p^{(3)} \ = \ & \tfrac{G_0 p}{N^2} \ \Big(\ \frac{2p-1}{4}K_3 - \frac{1}{4}(K_2)^2 - \frac{1}{2}K_2(x_i,x_j) \ \Big) \\
& + \tfrac{G_0}{N^2} \ \Big(\ \frac{3}{8}K_5 + \frac{1}{3}K_4 - \frac{1}{32}K_3^2 - \frac{1}{8}K_2K_1(x_i,x_j) \\
& \quad -\frac{1}{2}K_3(x_i,x_j) - \frac{1}{12}L_{1,1,1} - \frac{1}{2}L_{2,1,1} - \frac{1}{4}L_{2,2,1} \\
& \quad -\frac{1}{48}K_3L_{1,1,1} - \frac{1}{4}L_{3,1,1} + \frac{1}{8}K_1(x_i,x_j)^2 + \frac{1}{2}Q_1(x_i,x_j,x_k) \\
& \quad +\frac{1}{4}Q_2(x_i,x_j) + \frac{1}{2}Q_3(x_i,x_j,x_k) + \frac{1}{4}Q_4(x_i,x_j)
\end{aligned}
$$

$$+\frac{1}{6}L_{1,1,1}(x_i,x_j,x_k)+\frac{1}{24}L_{1,1,1}K_1(x_i,x_j)+\frac{1}{288}L^2_{1,1,1}$$
$$+\frac{1}{48}M_1+\frac{1}{8}M_2+\frac{1}{24}M_3+\frac{1}{16}M_4)$$

where

$$K_{a,b,\ldots}\equiv\sum_{1,2,\ldots}\rho_1^a\rho_2^b\ldots \tag{B.11}$$

$$K_a(x_i,x_j)\equiv\sum_{i,j}{}'\sum_1\rho_1^a e_{\vec{n}_1}(x_i)e^*_{\vec{n}}(x_j) \tag{B.12}$$

$$L_{a,b,c}\equiv\sum_{1,2,3}\rho_1^a\rho_2^b\rho_3^c\delta_{1+2+3} \tag{B.13}$$

$$Q_1(x_i,x_j,x_k)\equiv\sum_{i,j,k}{}'\sum_{1,2}\rho_1\rho_2 e_{\vec{n}_1}(x_i)e^*_{\vec{n}_1}(x_j)e_{\vec{n}_2}(x_j)e^*_{\vec{n}_2}(x_k) \tag{B.14}$$

$$Q_2(x_i,x_j)\equiv\sum_{i,j}{}'\sum_{1,2}\rho_1\rho_2 e_{\vec{n}_1}(x_i)e^*_{\vec{n}_1}(x_j)e^*_{\vec{n}_2}(x_i)e_{\vec{n}_2}(x_i) \tag{B.15}$$

$$Q_3(x_i,x_j,x_k)\equiv\sum_{i,j,k}{}^{i}\sum_{1,2,3}\rho_1\rho_2 e_{\vec{n}_1}(x_i)e^*_{\vec{n}_1}(x_j)e_{\vec{n}_2}(x_j)e^*_{\vec{n}_2}(x_k)e_{\vec{n}_3}(x_j)e^*_{\vec{n}_3}(x_k) \tag{B.16}$$

$$Q_4(x_i,x_j)\equiv\sum_{i,j}\sum_{1,2}\rho_1^2\rho_{1+2}e_{\vec{n}_1}(x_i)e^*_{\vec{n}_1}(x_j) \tag{B.17}$$

$$L_{a,b,c}(x_i,x_j,x_k)\equiv\sum_{i,j,k}{}'\sum_{1,2,3}\rho_1^a\rho_2^b\rho_3^c e_{\vec{n}_1}(\vec{x}_i)e_{\vec{n}_2}(\vec{x}_j)e_{\vec{n}_3}(\vec{x}_k)\delta_{1+2+3} \tag{B.18}$$

$$Q_{a,b}(x_i,x_j,x_k)\equiv\sum_{i,j,k}{}'\sum_{1,2}\rho_1^a\rho_2^b e_{\vec{n}_1}(\vec{x}_i)e^*_{\vec{n}_1}(\vec{x}_j)e_{\vec{n}_2}(\vec{x}_j)e^*_{\vec{n}_2}(\vec{x}_k) \tag{B.19}$$

$$M_1\equiv\sum_{1,2,3,4}\rho_1\rho_2\rho_3\rho_4\delta_{1+2+3+4} \tag{B.20}$$

$$M_2\equiv\sum_{1,2,3,4,5}\rho_1\rho_2\rho_3\rho_4\rho_5\delta_{1+2-5}\delta_{3+4-1-2}\delta_{5-3-4} \tag{B.21}$$

$$M_3\equiv\sum_{1,2,3,4,5,6}\rho_1\rho_2\rho_3\rho_4\rho_5\rho_6\delta_{1-2-5}\delta_{2-3-6}\delta_{3-1-4}\delta_{4+5+6} \tag{B.22}$$

$$M_4\equiv\sum_{1,2,3,4,5,6}\rho_1\rho_2\rho_3\rho_4\rho_5\rho_6\delta_{1-2-5}\delta_{2-3-6}\delta_{3+6-4}\delta_{4-5-1} \tag{B.23}$$

with

$$\rho_i\equiv\frac{2z\sigma^2_{\vec{n}_i}}{1-2z\sigma^2_{\vec{n}_i}} \tag{B.24}$$

and $\sum^i_{i,j,k,\ldots}=\sum_{\vec{x}_i\neq\vec{x}_j\neq\vec{x}_k\ldots}$, $\sum_{1,2,\ldots}\equiv\sum_{\vec{n}_1,\vec{n}_2,\ldots}$ and $\delta_{1+8-2+\ldots}\equiv\delta(\vec{n}_1+\vec{n}_8-\vec{n}_2+\ldots)$.

Bibliography

[1] Douglas Adams. *The Hitchhiker's Guide to the Galaxy.* Pan London.

[2] E. N. Argyres, Ronald H. P. Kleiss, and Costas G. Papadopoulos. Amplitude estimates for multi - higgs production at high- energies. *Nucl. Phys.*, B391:42–56, 1993.

[3] E. N. Argyres, A. F. W. van Hameren, R. H. P. Kleiss, and C. G. Papadopoulos. Zero-dimensional field theory. *Eur. Phys. J.*, C19:567–582, 2001.

[4] S. D. Badger, E. W. N. Glover, and Valentin V. Khoze. Mhv rules for higgs plus multi-parton amplitudes. *JHEP*, 03:023, 2005.

[5] Frits A. Berends and W. Giele. The six gluon process as an example of weyl-van der waerden spinor calculus. *Nucl. Phys.*, B294:700, 1987.

[6] Frits A. Berends and W. T. Giele. Recursive calculations for processes with n gluons. *Nucl. Phys.*, B306:759, 1988.

[7] Frits A. Berends and W. T. Giele. Recursive calculations for processes with n gluons. *Nucl. Phys.*, B306:759, 1988.

[8] Frits A. Berends, R. Pittau, and R. Kleiss. Excalibur: A monte carlo program to evaluate all four fermion processes at lep-200 and beyond. *Comput. Phys. Commun.*, 85:437–452, 1995.

[9] Carola F. Berger, Zvi Bern, Lance J. Dixon, Darren Forde, and David A. Kosower. Bootstrapping one-loop qcd amplitudes with general helicities. 2006.

[10] Zvi Bern, Lance J. Dixon, and David A. Kosower. On-shell recurrence relations for one-loop qcd amplitudes. *Phys. Rev.*, D71:105013, 2005.

[11] Zvi Bern, Darren Forde, David A. Kosower, and Pierpaolo Mastrolia. Twistor-inspired construction of electroweak vector boson currents. *Phys. Rev.*, D72:025006, 2005.

[12] Andreas Brandhuber, Bill J. Spence, and Gabriele Travaglini. One-loop gauge theory amplitudes in n = 4 super yang-mills from mhv vertices. *Nucl. Phys.*, B706:150–180, 2005.

[13] B. Bratley, P. Fox and H. Niederreiter. *AMC Transactions on Modeling and Computer Simulation 2*, 3:195.

[14] Ruth Britto, Freddy Cachazo, Bo Feng, and Edward Witten. Direct proof of tree-level recursion relation in yang-mills theory. *Phys. Rev. Lett.*, 94:181602, 2005.

[15] Freddy Cachazo, Peter Svrcek, and Edward Witten. Twistor space structure of one-loop amplitudes in gauge theory. *JHEP*, 10:074, 2004.

[16] Francesco Caravaglios and Mauro Moretti. An algorithm to compute born scattering amplitudes without feynman graphs. *Phys. Lett.*, B358:332–338, 1995.

[17] Elie Cartan. *The Theory of Spinors*. Hermann.

[18] G. Corcella et al. Herwig 6.5 release note. 2002.

[19] P. De Causmaecker, R. Gastmans, W. Troost, and Tai Tsun Wu. Helicity amplitudes for massless qed. *Phys. Lett.*, B105:215, 1981.

[20] Daniel de Florian and Jose Zurita. Seven parton amplitudes from recursion relations. 2006.

[21] Michael Dinsdale, Marko Ternick, and Stefan Weinzierl. A comparison of efficient methods for the computation of born gluon amplitudes. *JHEP*, 03:056, 2006.

[22] Lance J. Dixon. Calculating scattering amplitudes efficiently. 1996.

[23] Lance J. Dixon. Twistor string theory and qcd. *PoS*, HEP2005:405, 2006.

[24] Lance J. Dixon, E. W. N. Glover, and Valentin V. Khoze. Mhv rules for higgs plus multi-gluon amplitudes. *JHEP*, 12:015, 2004.

[25] P. D. Draggiotis and R. Kleiss. Counting tree diagrams: Asymptotic results for qcd-like theories. *Eur. Phys. J.*, C23:701–706, 2002.

[26] Petros Draggiotis and Ronald Kleiss. Vertex counting: Statistical distribution of vertices in large sets of tree diagrams. *Eur. Phys. J.*, C27:291–296, 2003.

[27] Petros Draggiotis, Ronald H. P. Kleiss, and Costas G. Papadopoulos. On the computation of multigluon amplitudes. *Phys. Lett.*, B439:157–164, 1998.

[28] Petros D. Draggiotis. *Exploding QCD*. PhD thesis, 2003.

[29] Petros D. Draggiotis, Ronald H. P. Kleiss, and Costas G. Papadopoulos. Multi-jet production in hadron collisions. *Eur. Phys. J.*, C24:447–458, 2002.

[30] James H. Ettle and Tim R. Morris. Structure of the mhv-rules lagrangian. 2006.

[31] H. Faure. Discrepances de suites associees a un systeme de numeration (en dimension s). *Acta Arith.*, 41:337.

[32] George Georgiou, E. W. N. Glover, and Valentin V. Khoze. Non-mhv tree amplitudes in gauge theory. *JHEP*, 07:048, 2004.

[33] George Georgiou and Valentin V. Khoze. Tree amplitudes in gauge theory as scalar mhv diagrams. *JHEP*, 05:070, 2004.

[34] J. Halton. *Num.Math.*, 2:84.

[35] Jiri Hoogland and Ronald Kleiss. Discrepancy-based error estimates for quasi-monte carlo. i: General formalism. 1996.

[36] Jiri Hoogland and Ronald Kleiss. Discrepancy-based error estimates for quasi-monte carlo. iii: Error distributions and central limits. *Comput. Phys. Commun.*, 101:21–30, 1997.

[37] C. Itzykson and J-B. Zuber. *Quantum Field Theory*. McGraw-Hill.

[38] F. James. Monte carlo theory and practice. *Rep. Prog. Phys.*, 43:1145, 1980.

[39] Aggeliki Kanaki and Costas G. Papadopoulos. Helac: A package to compute electroweak helicity amplitudes. *Comput. Phys. Commun.*, 132:306–315, 2000.

[40] Aggeliki Kanaki and Costas G. Papadopoulos. Helac: A package to compute electroweak helicity amplitudes. *Comput. Phys. Commun.*, 132:306–315, 2000.

[41] R. Kleiss and W. James Stirling. Spinor techniques for calculating p anti-p → w+- / z0 + jets. *Nucl. Phys.*, B262:235–262, 1985.

[42] D.E. Knuth. *The Art of Computer Programming*, volume 1. Addison-Wesley Professional.

[43] David A. Kosower. Next-to-maximal helicity violating amplitudes in gauge theory. *Phys. Rev.*, D71:045007, 2005.

[44] M. Luescher. *Comp.Phys.Comm.*, 79:100, 1994.

[45] Fabio Maltoni and Tim Stelzer. Madevent: Automatic event generation with madgraph. *JHEP*, 02:027, 2003.

[46] Michelangelo L. Mangano, Stephen J. Parke, and Zhan Xu. Duality and multi - gluon scattering. *Nucl. Phys.*, B298:653, 1988.

[47] Paul Mansfield. The lagrangian origin of mhv rules. 2005.

[48] N. Metropolis. The beginning of the monte carlo method. *Los Alamos Science*, 15:125, 1987.

[49] N. Metropolis and S. Ulam. The monte carlo method. *Journal of the American Statistical Association*, 44:335.

[50] H. Niederreiter. *J.Number Theory*, 30:51.

[51] A. Owen. Monte carlo extension of quasi monte carlo. *Winter Simulation Conference - IEEE Press.*

[52] K. J. Ozeren and W. J. Stirling. Scattering amplitudes with massive fermions using bcfw recursion. 2006.

[53] S.J. Parke and T.R. Taylor. An amplitude for n gluon scattering. *Phys. Revi. Lett.*, 56:2459, 1986.

[54] R.B.Pearson P.Cvitanovič, B.Lautrup. *Phys.Review D*, 18:1937, 1978.

[55] R.Kleiss P.D.Draggiotis and C.G.Papadopoulos. On the computation of multigluon amplitudes. *Phys.Lett.B*, 439:157.

[56] M. Peskin and D. Schroeder. *An Introduction to Quantum Field Theory.* Perseus Books.

[57] F. Cachazo R.Britto and B.Feng. New recursion relations for tree amplitudes of gluons. *Nucl.Phys.B*, 715:499.

[58] Kasper Risager. A direct proof of the csw rules. *JHEP*, 12:003, 2005.

[59] R.J. Rivers. *Path Integral Methods in Quantum Field Theory.* CUP.

[60] Ch. Schlier. *Comp.Phys.Comm.*, 159:23, 2004.

[61] W.M. Schmidt. *Acta Arith.*, 21:45.

[62] Vladimir A. Smirnov. Evaluating feynman integrals. *Springer Tracts Mod. Phys.*, 211:1–244, 2004.

[63] I.M. Sobol. The distribution of points in a cube and the approximate evaluation of integrals. *USSR Comp. Math. Math. Phys. 7*, 4:86.

[64] G. 't Hooft. Counting planar diagrams with various restrictions. *Nucl.Phys.B*, 538:389, 1999.

[65] S.Mrenna T.Sjostrand, L.Lonnblad and P.Skands. Pythia 6.3 physics and manual. *FERMILAB-PUB-03-457*, page 454, 2003.

[66] Diana Vaman and York-Peng Yao. Qcd recursion relations from the largest time equation. *JHEP*, 04:030, 2006.

[67] Steven Weinberg. *The Quantum Theory of Fields.* CUP.

[68] Stefan Weinzierl. Introduction to monte carlo methods. 2000.

[69] Edward Witten. Perturbative gauge theory as a string theory in twistor space. *Commun. Math. Phys.*, 252:189–258, 2004.

[70] Jun-Bao Wu and Chuan-Jie Zhu. Mhv vertices and fermionic scattering amplitudes in gauge theory with quarks and gluinos. *JHEP*, 09:063, 2004.

[71] Jun-Bao Wu and Chuan-Jie Zhu. Mhv vertices and scattering amplitudes in gauge theory. *JHEP*, 07:032, 2004.

[72] A. Zee. *Quantum Field Theory in a Nutshell.* Princeton University Press.

[73] Chuan-Jie Zhu. The googly amplitudes in gauge theory. *JHEP*, 04:032, 2004.

Index

Summary

In this thesis my work on three quite different problems within the field of multi-particle perturbative calculations is presented. The main issues approached here are: the computational complexity of NLO calculations, Monte Carlo and Quasi-Monte Carlo integration methods and error estimates therein, and the progress in the level of analytical understanding of QCD-like theories using the recently advocated on-shell recursion relations.

In chapter 2 methods for counting Feynman diagrams with loops for theories with bosonic particles and vertices with an arbitrary amount of legs are presented. Feynman diagrams are counted with and without symmetry factors and asymptotic results for one and two loops are given. The computational complexity of the Caravaglios-Moretti algorithm in the loop level with effective vertices is cited, and a comparison of that algorithm with the ordinary diagram per diagram approach is presented.

A method for calculating the error in the Quasi-Monte Carlo method for numerical integration is presented in chapter 3. An estimator for the error is suggested there, based on the underlying assumption that the Quasi point-set is a typical member of the ensemble of similar point-sets that share the property of having the same degree of uniformity, as measured by a particular Diaphony. The estimator is analyzed for a number of test-functions in various dimensions and is demonstrated to perform better than the ordinary, pseudo-random estimator. The latter is shown to consistently underestimate the error, as expected. A number of modifications are discussed that can be used to reduce the relevant computational.

In chapter 4 a fully diagrammatic proof of a recently advocated on-shell recursion relation for tree-level amplitudes involving gluons, is presented. The recurrence relation was first suggested by R. Britto, B. Feng, C. Cachazo and proven by these authors and E.Witten. It decomposes any on-shell n-point amplitude in on-shell amplitudes with a smaller number of legs and complexified momenta. The analysis presented here is based on a direct diagrammatic correspondence of Feynman diagrams with terms in a BCFW decomposition, employing a particular (but commonly seen) choice of axial vectors for the polarization vectors of the external gluons in the axial gauge. It is shown that, within this choice of axial vectors, the BCFW decomposition can be reorganized in terms of Feynman diagrams.

Samenvatting

In voorliggend proefschrift stel ik mijn werk op drie verschillende probleemgebieden voor binnen het vakgebied van perturbatieve berekeningen voor processen met meerdere deeltjes. De thema's die worden benaderd zijn, ten eerste, de computationele complexiteit van NLO-berekeningen, Monte Carlo en Quasi-Monte Carlo integratiemethoden en hun foutschattingen, en de vooruitgang in het analytisch begrijpen van QCD-achtige theorieën die recent voorgestane on-shell recursierelaties gebruiken.

In hoofdstuk 2 worden methodes voorgesteld om Feynmandiagrammen met lussen te tellen voor theorieën met bosondeeltjes en vertices met een willekeurig aantal benen. Feynmandiagrammen worden geteld met en zonder symmetriefactoren en asymptotsche resultaten voor één en twee lussen worden voorgesteld. Daarnaast wordt de computationele complexiteit van het Caravaglios-Moretti algoritme aangehaald in NLO met effectieve vertices en wordt een vergelijking gemaakt tussen dit algoritme en de normale diagram-per-diagram-benadering.

In hoofdstuk 3 wordt een methode voorgesteld voor het berekenen van de fout in de Quasi-Monte Carlo methode voor numerieke integratie. Er wordt een foutschatter voorgesteld op basis van de onderliggende aanname dat de Quasi-puntenset een typisch lid is van de verzameling gelijkaardige puntensets met dezelfde graad van uniformiteit zoals gemeten door een specifieke diafonie. Deze foutschatter wordt geanalyseerd voor een aantal testfuncties in verschillende dimensies. Er wordt aangetoond dat deze schatter betere resultaten geeft dan de normale pseudo-willekeurige schatter, die, zoals verwacht en hier aangetoond, de fout systematisch onderschat. Verder wordt in dit hoofdstuk een aantal wijzigingen besproken om de relevante computationele kost te verminderen.

In hoofdstuk 4 wordt een volledig diagrammatisch bewijs geleverd van een recent voorgestane on-shell recursierelatie voor leading-order amplitudes waarbij gluons betrokken zijn. Deze recursierelatie werd voor het eerst voorgesteld door R. Britto, B. Feng, C. Cachazo en samen met E.Witten bewezen. De methode houdt de decompositie in van elke on-shell n-punt amplitude in on-shell amplitudes met een kleiner aantal benen en gecomplexifieerde momenta. De analyse die hier wordt gegeven is gebaseerd op een directe, diagrammatische overeenkomst tussen Feynmandiagrammen en termen in een BCFW decompositie. De analyse maakt gebruik van een specifieke (maar algemeen verspreide) keuze van axiale vectoren voor de polarisatievectoren van externe gluons in de axiale ijk. Er wordt aangetoond dat binnen deze keuze van axiale vectoren de BCFW-decompositie kan worden gereorganiseerd via Feymandiagrammen.

Curriculum Vitae

Achillefs Lazopoulos was born in Athens, on 13 November 1975. He studied physics at the University of Athens and earned his Master's in Science degree at the Center for Particle Theory, University of Durham. He has spent the last four years as a Ph.D. candidate at the Department of High Energy Physics, University of Nijmegen. He is married to Heleen Derveaux and is the proud father of Chiara .

Publications

The research material presented in this thesis has been published in the following articles

- Recursive equations for arbitrary scattering processes
  P. Draggiotis, A. van Hameren, R. Kleiss, A. Lazopoulos, C.G. Papadopoulos, M. Worek
  Published in Nucl.Phys.Proc.Suppl.160:255-260,2006. Also in *Eisenach 2006, Elementary particle theory* 255-260
  e-Print Archive: hep-ph/0607034

- Diagrammatic proof of the BCFW recursion relation for gluon amplitudes in QCD
  Petros D. Draggiotis, Ronald H.P. Kleiss, Achilleas Lazopoulos, Costas G. Papadopoulos
  Published in Eur.Phys.J.C46:741,2006
  e-Print Archive: hep-ph/0511288

- Error in Monte Carlo, quasi-error in quasi-Monte Carlo
  Ronald H. Kleiss, Achilleas Lazopoulos
  Published in Comput.Phys.Commun.175:93-115,2006
  e-Print Archive: hep-ph/0504085

- Error estimates in Monte Carlo and quasi-Monte Carlo integration
  A. Lazopoulos
  Published in Acta Phys.Polon.B35:2617-2632,2004

- Counting loop diagrams: Computational complexity of higher-order amplitude evaluation
  Ernst van Eijk, Ronald Kleiss, Achilleas Lazopoulos
  Published in Eur.Phys.J.C36:459-470,2004
  e-Print Archive: hep-ph/0406025

Acknowledgment

I would like to express my gratitude to my supervisor, prof. Ronald Kleiss for showing me how science is done, for working with me in every step of the research relevant to this thesis, as well as, for his support and understanding during the rather long period I have been commuting between my desk in Nijmegen and my family in Antwerpen.

Special thanks to my co-promotor, dr.Costas Papadopoulos, for his scientific guidance and his hospitality in N.C.S.R. Demokritos.

Many thanks to everybody in the department of High Energy Physics at Nijmegen, and particularly to dr. Wim Beenaker who suffered my everyday presence for four years in our common office, dr. Petros Draggiotis who showed me my way around in and out of the institute at Nijmegen, dr. Andre van Hameren with whom we had various stimulating conversations both within and outside the realms of physics, dr. Chris Dams for his assistance in my frequent and obscure (to me, not to him) computer related problems, and last but not least, our frequent visitor in the high energy physics department, prof. Ernestos Argyres for conversations that offered me a glance upon his deep and historical perspective on the field.

Life in Nijmegen has been a joy-ride thanks to my friends there, dr. Avi Arambatzis, Vangelis Chantzis and Dimitris Karavidas, whom I would like to thank for all those endless coffee or beer sessions accompanied by exhausting conversations on anything, during the long, gloomy, rainy nights of my Netherlands years.

One can influence other people much deeper than one realizes. I would like to thank, at the end of this adventure, those that, knowing it or not, brought me to its beginning, my teachers in science, Andreas Kassetas, Alexis Simopoulos (who has managed to be much more than a teacher to me), prof.C.Ktorides and prof.N.Glover.

It would be a rather long task to thank all my friends and relatives that have in one way or another supported me through the years, but I can't end this acknowledgement note without mentioning three people that have shaped this project as well as my whole life, with their love, their unconditional support and their ubiquitous presence, and to whom this thesis is dedicated: my grandmother Alexandra, my father George, and my wife Heleen.

P.S. Though their contribution in this thesis is less than crucial, I would like to mention two more people because their contribution in my life is immense: my brother, Thanos (who should also get all credit for the cover), and our daughter, Chiara (who provides us with smiles of undescribable shine).